

QY 241 RTLNSTSPFPEKLEAEATKTLILSLGTLAOLDFSMRSMDSMDROEVLINNVSSSSSTOI 300
 Db 241 RTLNSTSPFPEKLEAEATKTLILSLGTLAOLDFSMRSMDSMDROEVLINNVSSSSSTOI 300
 QY 301 YQASRIYCGHPGEGGGLKIKSLNWEEDNNYKALFGNGTEDEAETFDYNDSTPYCNDLAK 360
 Db 301 YQASRIYCGHPGEGGGLKIKSLNWEEDNNYKALFGNGTEDEAETFDYNDSTPYCNDLAK 360
 QY 361 NLESPLSRIIMKALPLLVGKILYTPDTPATROVMAEVNKTFOELAVFHDLSGMEEELS 420
 Db 361 NLESPLSRIIMKALPLLVGKILYTPDTPATROVMAEVNKTFOELAVFHDLSGMEEELS 420
 QY 421 PKITWPMESQMDLVRLMLDSRDHFWEOQLDMDTADODIYAFLAKHEDVOSSNGS 480
 Db 421 PKITWPMESQMDLVRLMLDSRDHFWEOQLDMDTADODIYAFLAKHEDVOSSNGS 480
 QY 481 VYTWRAFNETNOAARTISREMECVNLKLEPIATEVWLINXSMELDERKFMAGIYFTG 540
 Db 481 VYTWRAFNETNOAARTISREMECVNLKLEPIATEVWLINXSMELDERKFMAGIYFTG 540
 QY 541 ITPGSIELPHNYKKTIRADIDNVERTNKIKDGYMDGPRADEEMRYVWGAFALDQV 600
 Db 541 ITPGSIELPHNYKKTIRADIDNVERTNKIKDGYMDGPRADEEMRYVWGAFALDQV 600
 QY 601 EOAIIIRLVTEKKTGVYMOQMRPCYVDIDFLRVMSRSMPLFMTLAMIYSAVIKIGIV 660
 Db 601 EOAIIIRLVTEKKTGVYMOQMRPCYVDIDFLRVMSRSMPLFMTLAMIYSAVIKIGIV 660
 QY 661 YEKARLEKTRIMIGLDSNLTMSFSSILPLVLSAGLLVYLKGNLPLSDPSVAV 720
 Db 661 YEKARLEKTRIMIGLDSNLTMSFSSILPLVLSAGLLVYLKGNLPLSDPSVAV 720
 QY 721 FLVSFAVVTIIOGCLISTELFSRANLAAACGIIYFTLYLPVLCVAMODYGFTLKIEAS 780
 Db 721 FLVSFAVVTIIOGCLISTELFSRANLAAACGIIYFTLYLPVLCVAMODYGFTLKIEAS 780
 QY 781 LLSVAFEGCEYFALFEBOGIGYOMDNLEESVYEDGFLTISISMILFDFLYGVMT 840
 Db 781 LLSVAFEGCEYFALFEBOGIGYOMDNLEESVYEDGFLTISISMILFDFLYGVMT 840
 QY 841 YIEAFPCOYIGIPRWYPCPKSYWFEESDEKSHPSNOKRMEICOMESEPTHLKIGVS 900
 Db 841 YIEAFPCOYIGIPRWYPCPKSYWFEESDEKSHPSNOKRMEICOMESEPTHLKIGVS 900
 QY 901 IONLVKVRDGMKAVDGLALNTEGOTTSFLGNGAGKTIMSILTLGPPPSGATYL 960
 Db 901 IONLVKVRDGMKAVDGLALNTEGOTTSFLGNGAGKTIMSILTLGPPPSGATYL 960
 QY 961 GMDISEKMTIPONLGYCPOHNVLFMDLTVEHIMFYARLKGLESEKHYKAEEMOMALDVG 1020
 Db 961 GMDISEKMTIPONLGYCPOHNVLFMDLTVEHIMFYARLKGLESEKHYKAEEMOMALDVG 1020
 QY 1021 LPSKSLKSTQSLSGMOKRISVALAFVGSKVYLLDEPTAGVDPYSRGIIWELLKXRO 1080
 Db 1021 LPSKSLKSTQSLSGMOKRISVALAFVGSKVYLLDEPTAGVDPYSRGIIWELLKXRO 1080
 QY 1081 GRITIIISTHMBADVLGRIRAIISHGKLCYVSSLEFLKNOLGTYITLVKKNVSSLS 1140
 Db 1081 GRITIIISTHMBADVLGRIRAIISHGKLCYVSSLEFLKNOLGTYITLVKKNVSSLS 1140
 QY 1141 SCNRSSSTVYLKEDSVSSOSSDAGLSGDSHSDTLTIDVSAISNLIRKHVSEALVEDI 1200
 Db 1141 SCNRSSSTVYLKEDSVSSOSSDAGLSGDSHSDTLTIDVSAISNLIRKHVSEALVEDI 1200
 QY 1201 GHELTYVLPYEAKEGAFVLLFHEIDDRISDLSISYISSETTLEPIFLKVAEESGVDAE 1260
 Db 1201 GHELTYVLPYEAKEGAFVLLFHEIDDRISDLSISYISSETTLEPIFLKVAEESGVDAE 1260
 QY 1261 TSOGTLPARNRRAFGQKQCLRPTEDDAADPNDSIDPESHETDLSGMDGKSYOVK 1320
 Db 1261 TSOGTLPARNRRAFGQKQCLRPTEDDAADPNDSIDPESHETDLSGMDGKSYOVK 1320
 QY 1321 GSKLTQOOFVALLMKRLILARRSKGFFAOIYLPANFVCIALVFSLIVPFQGYPSLEIQ 1380

Db 1381 GSKLTQOOFVALLMKRLILARRSKGFFAOIYLPANFVCIALVFSLIVPFQGYPSLEIQ 1380
 QY 1381 GSKLTQOOFVALLMKRLILARRSKGFFAOIYLPANFVCIALVFSLIVPFQGYPSLEIQ 1380
 Db 1381 GSKLTQOOFVALLMKRLILARRSKGFFAOIYLPANFVCIALVFSLIVPFQGYPSLEIQ 1380
 QY 1441 VPOTMDLPONGMTMOPSPACOCSSDKIKKMLPVCPPGAGLPPPOKQNTADILDL 1500
 Db 1441 VPOTMDLPONGMTMOPSPACOCSSDKIKKMLPVCPPGAGLPPPOKQNTADILDL 1500
 QY 1501 TGRNSTDVLTATYVOIILAKSLKNIWNEFRYGFSLGVSNTQALPPSOEVNDIKOMK 1560
 Db 1501 TGRNSTDVLTATYVOIILAKSLKNIWNEFRYGFSLGVSNTQALPPSOEVNDIKOMK 1560
 QY 1561 HLKAKSSADRFILSLGRFMTGLDTRNNKRVWPNKGMHAISSFLAVINNALIRANLQK 1620
 Db 1561 HLKAKSSADRFILSLGRFMTGLDTRNNKRVWPNKGMHAISSFLAVINNALIRANLQK 1620
 QY 1621 GNPSSHGTTAFNHPNLNKOOLSEVALMTTSVDVLSICVIFAMSEVPASVYVFLIOER 1680
 Db 1621 GNPSSHGTTAFNHPNLNKOOLSEVALMTTSVDVLSICVIFAMSEVPASVYVFLIOER 1680
 QY 1681 VSKAKHLOFISGVKPYIYWLNSFVDMCNVYPATLVIIITICQOKSYSSSTLPLVAL 1740
 Db 1681 VSKAKHLOFISGVKPYIYWLNSFVDMCNVYPATLVIIITICQOKSYSSSTLPLVAL 1740
 QY 1741 LLLVGMSTPLMPASFEFKIPSTAYVLTSVNLFIGINSVATFVLELTDKLNIN 1800
 Db 1741 LLLVGMSTPLMPASFEFKIPSTAYVLTSVNLFIGINSVATFVLELTDKLNIN 1800
 QY 1801 DILKSVFLIPHFCLRGCLIDVKNQAMADDERFGENRFPVSLWDLVGRNLPMAVEG 1860
 Db 1801 DILKSVFLIPHFCLRGCLIDVKNQAMADDERFGENRFPVSLWDLVGRNLPMAVEG 1860
 QY 1861 VVFLITVLIOYRFTLRPRVNAKLSPLNDEDEDVREBORILIDGGQNDLEIKELTKI 1920
 Db 1861 VVFLITVLIOYRFTLRPRVNAKLSPLNDEDEDVREBORILIDGGQNDLEIKELTKI 1920
 QY 1921 YRRRRKPAVDRIICVIGPBGCGILGVNAGKSSFTKMLTGDTTYRGAFILNKSIISN 1980
 Db 1921 YRRRRKPAVDRIICVIGPBGCGILGVNAGKSSFTKMLTGDTTYRGAFILNKSIISN 1980
 QY 1981 IHEYHONMGYCPQDATTETLLTGREHEFEFALLRGVPEKVGKGMATRKILGLVYGEK 2040
 Db 1981 IHEYHONMGYCPQDATTETLLTGREHEFEFALLRGVPEKVGKGMATRKILGLVYGEK 2040
 QY 2041 YAGNYSGNKRLSTAMALIGPPVFLDEPTGMDKARRFILNICALSVYKBSRSVILT 2100
 Db 2041 YAGNYSGNKRLSTAMALIGPPVFLDEPTGMDKARRFILNICALSVYKBSRSVILT 2100
 QY 2101 SHSMECEALCTRAIMVNGFRCLGSVOHLKNRFGGYTYIVRIAGSNPDLKPYODFEG 2160
 Db 2101 SHSMECEALCTRAIMVNGFRCLGSVOHLKNRFGGYTYIVRIAGSNPDLKPYODFEG 2160
 QY 2161 LAFPGSVLEKRRNMLQYOLPSSLSLARIFSILSOSKRLIHIEDYSVOTTLDQVFN 2220
 Db 2161 LAFPGSVLEKRRNMLQYOLPSSLSLARIFSILSOSKRLIHIEDYSVOTTLDQVFN 2220
 QY 2221 AKDOSDDHLKDLSLHKNOTVVDVAVLTSFLIODEKVEKSY 2261
 Db 2221 AKDOSDDHLKDLSLHKNOTVVDVAVLTSFLIODEKVEKSY 2261

RESULT 2

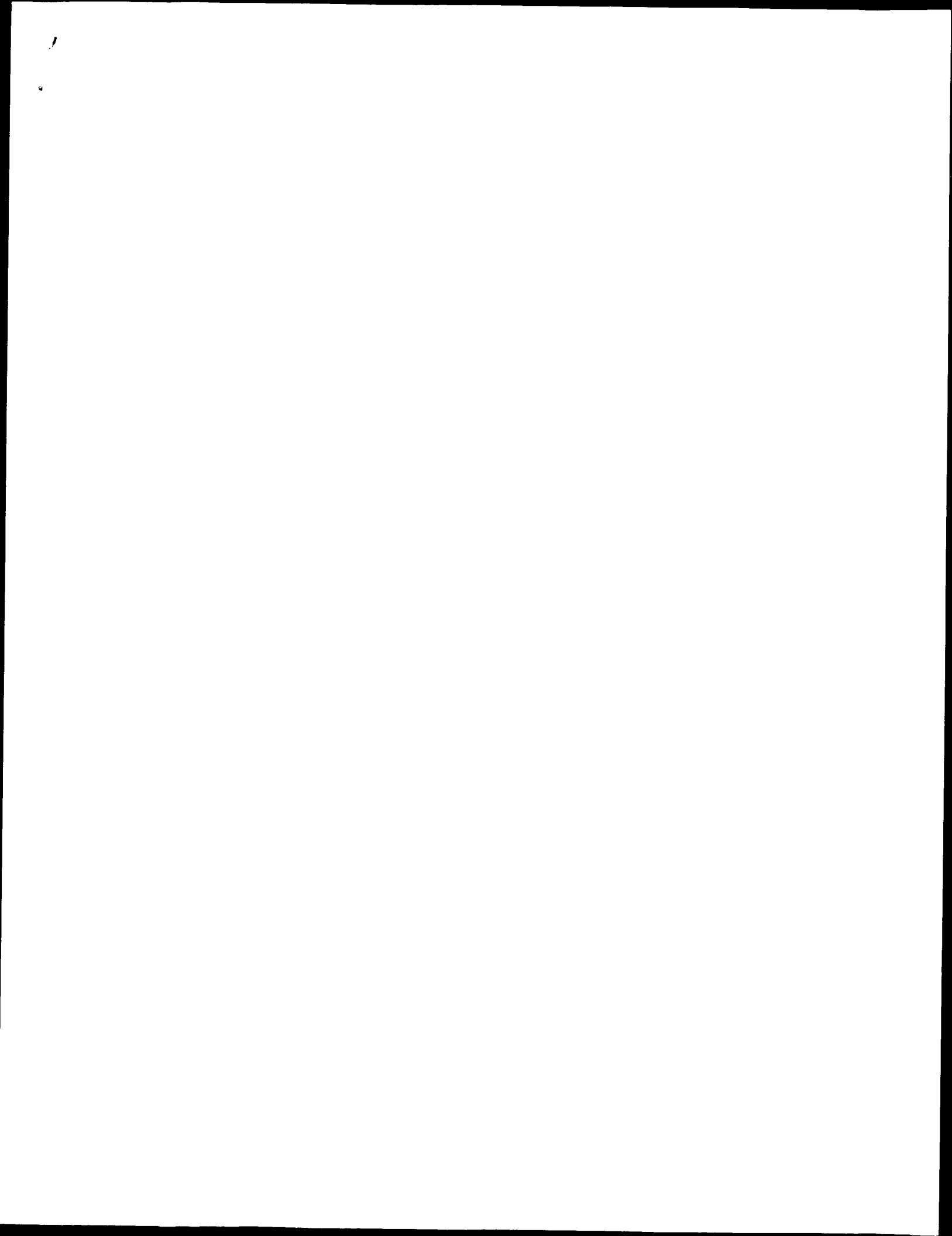
US-09-846-456-11
 ; Sequence 11, Application US/09846456
 ; Patent No. US20020146792A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Rosier, Marie
 ; APPLICANT: Prades, Catherine
 ; APPLICANT: Lemoine, Cendrine
 ; APPLICANT: Naudin, Laurent

APPLICANT: Deneffe, Patrice
 APPLICANT: Duverger, Nicolas
 APPLICANT: Brewer, Brian
 APPLICANT: Remaley, Alan
 APPLICANT: Fojo, Sylvia
 TITLE OF INVENTION: Regulatory Nucleic Acid for the ABC1 Gene, Molecules Modifying It
 TITLE OF INVENTION: Activity and Therapeutic Uses
 FILE REFERENCE: 3806.0505
 CURRENT APPLICATION NUMBER: US/09/846,456
 CURRENT FILING DATE: 2001-05-02
 PRIOR APPLICATION NUMBER: US 60/201,280
 PRIOR FILING DATE: 2000-05-02
 NUMBER OF SEQ ID NOS: 20
 SOFTWARE: PatentIn version 3.0
 SEQ ID NO 11
 LENGTH: 2261
 TYPE: PRP
 ORGANISM: Homo sapiens
 US-09-846-456-11

Query Match 99.7%; Score 11767; DB 10; Length 2261;
 Best Local Similarity 99.7%; Pred. No. 0;
 Matches 2254; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 MACPOLRLMLKNTLFRFRROTQQLLEVAWPLFILLISVRLSPYEOHECHPNKA 60
 DB 1 MACPOLRLMLKNTLFRFRROTQQLLEVAWPLFILLISVRLSPYEOHECHPNKA 60
 QY 61 MPAGTLPMVOGIIICNANNECFRYPTEGAPGVGNFNKSIARLFSRARLLLYSQKDT 120
 DB 61 MPAGTLPMVOGIIICNANNECFRYPTEGAPGVGNFNKSIARLFSRARLLLYSQKDT 120
 QY 121 SMKDMKRLTLOQIKSSSNLKLQDFLVNDFEFGFLVHNLSPKSTYDKMLRAVILH 180
 DB 121 SMKDMKRLTLOQIKSSSNLKLQDFLVNDFEFGFLVHNLSPKSTYDKMLRAVILH 180
 QY 181 KYVLOGYQLHLTSLCNGSKSEEMTOLGDOVESELGKPKKLAARVLRNSNDILKPTL 240
 DB 181 KYVLOGYQLHLTSLCNGSKSEEMTOLGDOVESELGKPKKLAARVLRNSNDILKPTL 240
 QY 241 RTINSTSPFSKBLAETATKTLHSLSGLTQAELEFSMSRWSMDROEVMFLTNVSSSSTQI 300
 DB 241 RTINSTSPFSKBLAETATKTLHSLSGLTQAELEFSMSRWSMDROEVMFLTNVSSSSTQI 300
 QY 301 YQAVSRIVCGHPREGGGLKISLWYEDNNKALFGNGTEDEAETTYDNTSTPYCNDLAK 360
 DB 301 YQAVSRIVCGHPREGGGLKISLWYEDNNKALFGNGTEDEAETTYDNTSTPYCNDLAK 360
 QY 361 NLESSPLSRITMKALKPLVGLIYTPDPATROVAEVAKTFQELAVFHDLEGMWELS 420
 DB 361 NLESSPLSRITMKALKPLVGLIYTPDPATROVAEVAKTFQELAVFHDLEGMWELS 420
 QY 421 PKIWTFMENSQENDLVMLDSDRNDHFWEQQLDGLDWTADIVAFIADAKREDOVQSSNGS 480
 DB 421 PKIWTFMENSQENDLVMLDSDRNDHFWEQQLDGLDWTADIVAFIADAKREDOVQSSNGS 480
 QY 481 VYTWREAFNETNOAIRTISRMECVNLKLEPIATEYVWLINKSMELLDEKKFAGIYFTG 540
 DB 481 VYTWREAFNETNOAIRTISRMECVNLKLEPIATEYVWLINKSMELLDEKKFAGIYFTG 540
 QY 541 ITPGSIELPHHVKKIRMDIDNVERTNKIKDGYWDPGRADPEDMRYVWGFAIYLDVY 600
 DB 541 ITPGSIELPHHVKKIRMDIDNVERTNKIKDGYWDPGRADPEDMRYVWGFAIYLDVY 600
 QY 601 EOLATIRVLGTEKKTGYMOOMPVCYVDIIFLRVMSRSPLEMTLAWISVAVIIGIY 660
 DB 601 EOLATIRVLGTEKKTGYMOOMPVCYVDIIFLRVMSRSPLEMTLAWISVAVIIGIY 660
 QY 661 YEKEARLKETIRKINGLDSILWFSWISLIPPLVAGLLVILKGNLPLSPDPVVEV 720
 DB 661 YEKEARLKETIRKINGLDSILWFSWISLIPPLVAGLLVILKGNLPLSPDPVVEV 720
 QY 721 FLVSFAVYVTTIQCFLISTLSRANLAACGGIYFTLYLPYVLCVAMQDYVGFTLKIFAS 780

DB 721 FLVSFAVYVTTIQCFLISTLSRANLAACGGIYFTLYLPYVLCVAMQDYVGFTLKIFAS 780
 QY 781 LLSVAFEGGCEYFALFEEOGIGVQWNLNLFESPEVEDGNLTTSISMLFDTLGYVMW 840
 DB 781 LLSVAFEGGCEYFALFEEOGIGVQWNLNLFESPEVEDGNLTTSISMLFDTLGYVMW 840
 QY 841 YIEAVFGQYIIPRPWYFPCTKSYWFEEDSKSHPSNKRSEICMEEPHKLKIGVS 900
 DB 841 YIEAVFGQYIIPRPWYFPCTKSYWFEEDSKSHPSNKRSEICMEEPHKLKIGVS 900
 QY 901 IONLVKVRDGMKVAVQGLALNFEQOITSPFGNAGKTTMSILTGLPPTSGTYIL 960
 DB 901 IONLVKVRDGMKVAVQGLALNFEQOITSPFGNAGKTTMSILTGLPPTSGTYIL 960
 QY 961 GKDIRSEMTIRONLGYCPOHNVLFDMLTVEEHIWYARLKGISEKHVAKMEQMALDVG 1020
 DB 961 GKDIRSEMTIRONLGYCPOHNVLFDMLTVEEHIWYARLKGISEKHVAKMEQMALDVG 1020
 QY 1021 LPSSKLSKTSQLSGGMQRKLSVALAFVGSKYVILDEPTAGVDPYRRGIWELLKYRQ 1080
 DB 1021 LPSSKLSKTSQLSGGMQRKLSVALAFVGSKYVILDEPTAGVDPYRRGIWELLKYRQ 1080
 QY 1081 GRTIILSTHMDADVLDRIAIISHGKLCVGSLSFLKNOIGTYLTVKKDVESSL 1140
 DB 1081 GRTIILSTHMDADVLDRIAIISHGKLCVGSLSFLKNOIGTYLTVKKDVESSL 1140
 QY 1141 SCRNSSSTVSTLKEDSVSOSSDAGLSHESOTLITVSAISNLIRKHVSEARLYEDI 1200
 DB 1141 SCRNSSSTVSTLKEDSVSOSSDAGLSHESOTLITVSAISNLIRKHVSEARLYEDI 1200
 QY 1201 GHELTYYVPEAKRGAVELFHEIDRLSDLGTSYGISTTLEELFLKAAEESGDAE 1260
 DB 1201 GHELTYYVPEAKRGAVELFHEIDRLSDLGTSYGISTTLEELFLKAAEESGDAE 1260
 QY 1261 TSDGTLPARNRNRAFQKOSCLPFTEDDADPNDSDIDPESRETDLSCMDGKSGYQV 1320
 DB 1261 TSDGTLPARNRNRAFQKOSCLPFTEDDADPNDSDIDPESRETDLSCMDGKSGYQV 1320
 QY 1321 GSKLTQOQVALLMKRLIARSRKGFPAQIVLAVFVICALVSLVPPGKPSLELQ 1380
 DB 1321 GSKLTQOQVALLMKRLIARSRKGFPAQIVLAVFVICALVSLVPPGKPSLELQ 1380
 QY 1381 PMYNEQYTFVSNDAPEDTGLLELNLATKDPGFCRMBENPIPDPTCQAGEEMTAP 1440
 DB 1381 PMYNEQYTFVSNDAPEDTGLLELNLATKDPGFCRMBENPIPDPTCQAGEEMTAP 1440
 QY 1441 VPOTIMDLFONGNMTNONSAPACQSSDKIKMLPVCPPGAGLPPQKONTADILQDL 1500
 DB 1441 VPOTIMDLFONGNMTNONSAPACQSSDKIKMLPVCPPGAGLPPQKONTADILQDL 1500
 QY 1501 TGRNISDYLVKTYOIIAKSLKNKIWNEFRYGGSLVSNTOALPQOEVDATKQKK 1560
 DB 1501 TGRNISDYLVKTYOIIAKSLKNKIWNEFRYGGSLVSNTOALPQOEVDATKQKK 1560
 QY 1561 HLKAKDSSADRFNLISLGFMTGJLDRNNVYKWFNNKGMAHAISSFLYINNAIIRANLQK 1620
 DB 1561 HLKAKDSSADRFNLISLGFMTGJLDRNNVYKWFNNKGMAHAISSFLYINNAIIRANLQK 1620
 QY 1621 GENPSHYGTTAANHPNLTKOOLSEVALMTTSVDVLYSICVIFAMSFPASVYFLIOER 1680
 DB 1621 GENPSHYGTTAANHPNLTKOOLSEVALMTTSVDVLYSICVIFAMSFPASVYFLIOER 1680
 QY 1681 VSKAKHLOFISGVKPYIYLSNFWDMCNVYVPAIVLIIIFCFOOKSYVSTNLPVAL 1740
 DB 1681 VSKAKHLOFISGVKPYIYLSNFWDMCNVYVPAIVLIIIFCFOOKSYVSTNLPVAL 1740
 QY 1741 LLLLYGMSITPLMPASVFKIPSTAYVVLTSVNLFIGINGSVAFVLELFDNKLNNIN 1800
 DB 1741 LLLLYGMSITPLMPASVFKIPSTAYVVLTSVNLFIGINGSVAFVLELFDNKLNNIN 1800
 QY 1801 DILKSYFLIFPHFCUGRGLIDWYKQAMADALERGENRFVPSLWDLYGRULFAMAVEG 1860



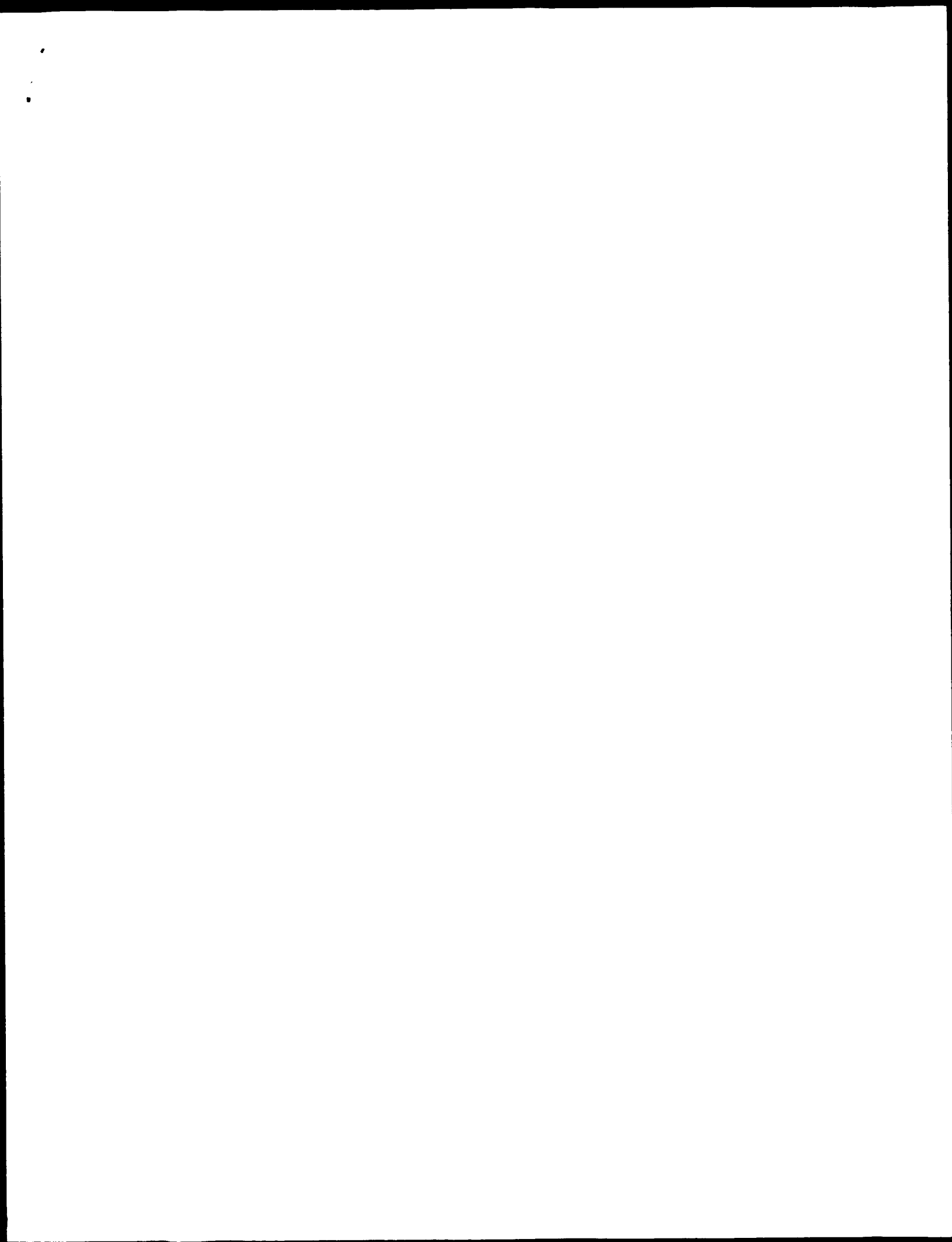
Db 1801 DLKSVFLIFPHFCLGSLIDMVKNAADALBERFGENRFSPLSMDYGRNLFAMAVEG 1860
 Qy 1861 VFEFLITVLIOYREFIRPPVNAKLSPLNDEDEVDREORRIIDGGGONDILEIKELTKY 1920
 Db 1861 VFEFLITVLIOYREFIRPPVNAKLSPLNDEDEVDREORRIIDGGGONDILEIKELTKY 1920
 Qy 1921 YRRKRRPAVDRIICVGIIPGBCFGLGVNAGKSGSTFKMLTGTVTYRGDAFLKNSILSN 1980
 Db 1921 YRRKRRPAVDRIICVGIIPGBCFGLGVNAGKSGSTFKMLTGTVTYRGDAFLKNSILSN 1980
 Qy 1981 IHEHOMMGVCPQPDATTELLTGREHEFFALLRGVPEKGVGEMATKLCVYKGEK 2040
 Db 1981 IHEHOMMGVCPQPDATTELLTGREHEFFALLRGVPEKGVGEMATKLCVYKGEK 2040
 Qy 2041 YAGNYSGNKRLKSTAMALIGPPVVELDEPTGMDPKARFLMNCALSVYKGRSVILT 2100
 Db 2041 YAGNYSGNKRLKSTAMALIGPPVVELDEPTGMDPKARFLMNCALSVYKGRSVILT 2100
 Qy 2101 SHSMECEALCTRMALVNGRFRCLGSVOHLKNNFGDGYTIVIRIAGSNEDLKPVODFG 2160
 Db 2101 SHSMECEALCTRMALVNGRFRCLGSVOHLKNNFGDGYTIVIRIAGSNEDLKPVODFG 2160
 Qy 2161 LAFPGSVPEKRRHNMLOLYOLPSSLSLARIFSLSSKRRLEEDYSVOTLLDOYVNF 2220
 Db 2161 LAFPGSVPEKRRHNMLOLYOLPSSLSLARIFSLSSKRRLEEDYSVOTLLDOYVNF 2220
 Qy 2221 AKDOSDDHLKDLHLKNTQTVAVVAVLTSFLODEKYESYV 2261
 Db 2221 AKDOSDDHLKDLHLKNTQTVAVVAVLTSFLODEKYESYV 2261

RESULT 3
 US-09-595-542-9
 ; Sequence 9, Application US/09995542
 ; Patent No. US20020187647A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Shutter, John
 ; APPLICANT: Ollas, Laetia
 ; TITLE OF INVENTION: ATP-Binding Cassette Transporter-Like Molecules and
 ; FILE REFERENCE: 00-658-A
 ; CURRENT APPLICATION NUMBER: US-09/995,542
 ; CURRENT FILING DATE: 2001-11-28
 ; PRIOR APPLICATION NUMBER: 60/253,520
 ; PRIOR FILING DATE: 2000-11-28
 ; NUMBER OF SEQ ID NOS: 24
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 9
 ; LENGTH: 2201
 ; TYPE: PRT
 ; ORGANISM: Mus musculus
 ; FEATURE:
 ; NAME/KEY: UNSURE
 ; LOCATION: (115)
 ; OTHER INFORMATION: amino acid at this position is unknown
 US-09-995-542-9

Query Match 92.6%; Score 10922; DB 10; Length 2201;
 Best Local Similarity 94.9%; Pred. No. 0;
 Matches 2089; Conservative 55; Mismatches 57; Indels 0; Gaps 0;

Qy 61 MPASGILPWVVOGIIICANNPCFRYPGEGAPGVGNFNKSTIVARLFSDAKRLILYSOKDT 120
 Db 1 MPASGILPWVVOGIIICANNPCFRYPGEGAPGVGNFNKSTIVARLFSDAKRLILYSOKDT 60
 Qy 121 SMDMRVLTLOQIKSSNLTLODFIVDNTEFSGFLVHNLSLPSTYDKMLRAVYLIL 180
 Db 61 SMDMRVLTLOQIKSSNLTLODFIVDNTEFSGFLVHNLSLPSTYDKMLRAVYLIL 180
 Qy 181 KVFLOGYQLHLTSLGSKSEEMIOIGDQVSESLGLPKKELAAARVRLRSNMDILKPII 240
 Db 121 KVFLOGYQLHLTSLGSKSEEMIOIGDQVSESLGLPKKELAAARVRLRSNMDILKPII 180

Qy 241 RTLNSTSPSPSKELAEATKTLHLSTGLAOLFMSRMSKSDMKROEMFLTNVSSSSSTOI 300
 Db 181 TLTNSTSHLPFOHLAEATTVLLDLSGLAOLFMSRMSKSDMKROEMFLTNVSSSSSTOI 240
 Qy 301 YQAVSRIYCGHEGGGLIKSLNMYEDNNYKALFGNGTEDEAETFYDNSTTPCNDLMK 360
 Db 241 YQAVSRIYCGHEGGGLIKSLNMYEDNNYKALFGNGTEDEAETFYDNSTTPCNDLMK 300
 Qy 361 NLESSPISRIYKALKPLLVKILYTPPTPATROYMAEVNKTFOELAVFHDLBGMWELS 420
 Db 301 NLESSPISRIYKALKPLLVKILYTPPTPATROYMAEVNKTFOELAVFHDLBGMWELS 360
 Qy 421 PKIWFEMESQMDIVRMLDSRDNDHFWEODDLDMTAADIYAFIAKHPEDVOSSGS 480
 Db 361 PKIWFEMESQMDIVRMLDSRDNDHFWEODDLDMTAADIYAFIAKHPEDVOSSGS 420
 Qy 481 VYTRDEAFNETNOALITTSREMECVNLKLEPIATEVWLINKSMELDERKFMAGIVFTG 540
 Db 421 VYTRDEAFNETNOALITTSREMECVNLKLEPIATEVWLINKSMELDERKFMAGIVFTG 480
 Qy 541 ITPGSLIELPHHKYKIRMDIDNVERTNKIKQGYNDPGPRADPFEDMKRYWVGCFAYLDYV 600
 Db 481 ITPGSLIELPHHKYKIRMDIDNVERTNKIKQGYNDPGPRADPFEDMKRYWVGCFAYLDYV 540
 Qy 601 EGAIRVLTGTEKKTGYVMQMPYCYVDJFELRVMRSMPLEPTLAWIYSAVIYIKGIV 660
 Db 541 EGAIRVLTGTEKKTGYVMQMPYCYVDJFELRVMRSMPLEPTLAWIYSAVIYIKGIV 600
 Qy 661 YEKEARKETMRIMGDLSILMFSWFTSSKIPLLVSAGLVVILKGNLIPYSDPSVAFV 720
 Db 601 YEKEARKETMRIMGDLSILMFSWFTSSKIPLLVSAGLVVILKGNLIPYSDPSVAFV 660
 Qy 721 FLTSFAVVTILOCFLISLFLSRANLAACGGITFETLYLPVYLAVANODYGFLTKIFAS 780
 Db 661 FLTSFAVVTILOCFLISLFLSRANLAACGGITFETLYLPVYLAVANODYGFLTKIFAS 720
 Qy 781 LLSPAVFGGCEYFALFEEOGIGVQMDNLFPSPVEEDPFNTITSIMKLEPTFLYGYMTV 840
 Db 721 LLSPAVFGGCEYFALFEEOGIGVQMDNLFPSPVEEDPFNTITSIMKLEPTFLYGYMTV 780
 Qy 841 YLEAVFPGQYIGIPRWYTPCTKSTWFGSEDEKSHPSGNOKRSEICMEEPFLKLGVS 900
 Db 781 YLEAVFPGQYIGIPRWYTPCTKSTWFGSEDEKSHPSGNOKRSEICMEEPFLKLGVS 840
 Qy 901 IONLVKVRDGMKAVVDLALNFEYGOITSFIGNHAGKTTMTSLSLPPTSGTAYIL 960
 Db 841 IONLVKVRDGMKAVVDLALNFEYGOITSFIGNHAGKTTMTSLSLPPTSGTAYIL 900
 Qy 961 GKDISEMSTIRONLGYCPQHNVLJEDMLTVEHIMFYARKLGSEKHVKAEMQALDVG 1020
 Db 901 GKDISEMSTIRONLGYCPQHNVLJEDMLTVEHIMFYARKLGSEKHVKAEMQALDVG 960
 Qy 1021 LPSSKLKSTQSLSGMKRKLISVALAFVGSKVYILDEPTAGVDPYSRGIWELLKYYRQ 1080
 Db 961 LPSSKLKSTQSLSGMKRKLISVALAFVGSKVYILDEPTAGVDPYSRGIWELLKYYRQ 1020
 Qy 1081 GRTIILSTHMDADYLDRIAIIISHGKLCVGSLSFLKNOLGYYILTVKKXVESSLS 1140
 Db 1021 GRTIILSTHMDADYLDRIAIIISHGKLCVGSLSFLKNOLGYYILTVKKXVESSLS 1080
 Qy 1141 SCRNSSSTVSYLKEDSVSOSSDAGLSDESDTLTIDVSAISNLIRKHVSEARLEDI 1200
 Db 1081 SCRNSSSTVSYLKEDSVSOSSDAGLSDESDTLTIDVSAISNLIRKHVSEARLEDI 1140
 Qy 1201 GHETLYVLPYEAKAGAVELFHEIDRLSDLGISSTGISTTLEETFLKVAEESGVDAE 1260
 Db 1141 GHETLYVLPYEAKAGAVELFHEIDRLSDLGISSTGISTTLEETFLKVAEESGVDAE 1200
 Qy 1261 TSDGTLPARRRNRAFGDQSCRPTEEDADPNDSDIDPESRPTDLISGMDGSGSYOVK 1320
 Db 1201 TSDGTLPARRRNRAFGDQSCRPTEEDADPNDSDIDPESRPTDLISGMDGSGSYOVK 1260
 Qy 1321 GWKLTQGFVALMKRLLIARRSRKGFPAQIVLPVAVFCIALVSLVPPGKXPSLEIQ 1380



GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 7, 2003, 08:41:19 ; Search time 41 Seconds
(without alignments)
2287.268 Million cell updates/sec

Title: US-09-595-526c-2

Perfect score: 11797
Sequence: 1 MACWQPLRLTLWKNTLFRRR.....VDVAVLTSFLDQEVKESYV 2261

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database: SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	11770	99.8	2261	1	ABCI_HUMAN
2	11256	95.4	2261	1	ABCI_MOUSE
3	5870.5	49.8	2273	1	ABCR_HUMAN
4	4230.5	35.9	2436	1	ABCR2_MOUSE
5	4093.5	34.7	2434	1	ABCR2_MOUSE
6	2642.5	22.4	1704	1	ABCR2_MOUSE
7	1538.5	13.0	1704	1	CDP7_CAERL
8	411	3.5	330	1	DRRA_STRE
9	382.5	3.2	343	1	NOD1_RHISN
10	366	3.1	304	1	NOD1_RHISN
11	347	2.9	308	1	NOD1_RHISN
12	344.5	2.9	340	1	NOD1_RHISN
13	343.5	2.9	347	1	NOD1_RHISN
14	339.5	2.9	1280	1	NOD1_HUMAN
15	331.5	2.8	355	1	NOD1_RHISN
16	329.5	2.8	578	1	YBHF_ECOLI
17	327.5	2.8	1281	1	MDR3_CRIGR
18	327	2.8	894	1	YHIF_ECOLI
19	326	2.8	1276	1	MDR3_MOUSE
20	325.5	2.8	308	1	NOSE_PSEST
21	324.5	2.8	1276	1	MDR2_MOUSE
22	322.5	2.7	306	1	NOD1_MOUSE
23	321	2.7	1276	1	MDR1_CRIGR
24	317	2.7	1362	1	PMD1_SCHPO
25	316	2.7	1276	1	MDR2_CRIGR
26	315.5	2.7	311	1	NOD1_RHISN
27	315	2.7	1278	1	MDR1_MOUSE
28	314.5	2.7	1277	1	MDR1_MOUSE
29	313	2.7	354	1	Y415_SYNY3
30	310.5	2.6	1276	1	MDR1_MOUSE
31	304.5	2.6	1294	1	YOH5_YEAST
32	300	2.5	381	1	OPB5_BACSU
33	298.5	2.5	262	1	YA23_METUA

34	297	2.5	380	1	OPCA_BACSU
35	296	2.5	335	1	V719_ANASP
36	293.5	2.5	1321	1	AB11_HUMAN
37	292.5	2.5	305	1	YHCH_BACSU
38	287.5	2.4	274	1	V179_MYCPN
39	286.5	2.4	274	1	V179_MYCPN
40	285.5	2.4	1279	1	MDR3_HUMAN
41	284	2.4	306	1	MDR3_HUMAN
42	284	2.4	1321	1	MDR3_HUMAN
43	284	2.4	1336	1	MDR3_HUMAN
44	281.5	2.4	308	1	YHCH_ECOLI
45	281.5	2.4	343	1	ABC_ECOLI

ALIGNMENTS

RESULT 1
ABCI_HUMAN STANDARD: PRT: 2261 AA.
AC 095477: 09UN08: 09UN07: 09UN06: 09UN04: 09UN09: 096T85: 096S56:
16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ATP-binding cassette, sub-family A, member 1 (ATP-binding cassette transporter 1) (ATP-binding cassette 1) (ABC-1) (Cholesterol efflux regulatory protein).
GN ABC1 OR ABC1 OR CERP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID:9606;
RN [1]
RP MEDLINE=2034509; PubMed=10884428;
RX MEDLINE=2034509; PubMed=10884428;
RA Santamarina-Fojo S., Peterson K.M., Knapper C.L., Qiu Y., Freeman L.A., Cheng J.-F., Osorio J., Remley A.T., Yang X.-P., Haendelschild C.C., Prades C., Chimenti G., Blackton E.F., Francois T.L., Duverger N., Rubin E.W., Koster M., Deneffe P., Fredrickson D.S., Brewer H.B. Jr.;
RA "Complete genomic sequence of the human ABCA1 gene: analysis of the human and mouse ATP-binding cassette A promoter.";
RT Proc. Natl. Acad. Sci. U.S.A. 97:7987-7992(2000). -544.
RL [2]
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Skin;
RA Schwartz K., Lawn R.M., Wade D.P.;
RT "ABCA1 gene expression and apoA-I-mediated cholesterol efflux are regulated by LXR.";
RT Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=2151004; PubMed=11352567;
RA Qiu Y., Cavellier L., Chiu S., Yang X., Rubin E., Cheng J.-F.;
RT "Human and mouse ABCA1 comparative sequencing and transgenesis studies: revealing novel regulatory sequences.";
RL Genomics 73:66-76(2001).
RN [4]
RP SEQUENCE FROM N.A.
RA Tanaka A.R., Abe-Dohmae S., Arakawa R., Sadanami K., Kidera A., Kioke N., Amachi T., Yokoyama S., Ueda K.;
RT "A new topological model of functional human ABCA1-signal peptide cleavage and glycosylation of a large extracellular domain.";
RN Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE OF 21-2261 FROM N.A.
RX MEDLINE=99194549; PubMed=10092505;
RA Langman G., Klucken J., Reil M., Liebisch G., Luciani M.F., Chimenti G., Kaminski W.E., Schmitz G.;
RT "Molecular cloning of the human ATP-binding cassette transporter 1 (ABCA1): evidence for sterol-dependent regulation in macrophages.";
RL Biochem. Biophys. Res. Commun. 257:29-33(1999).
RN [6]

RP SEQUENCE OF 21-2261 FROM N.A.
 RX MEDLINE=99364413; PubMed=10431238;
 RA Rust S., Rosier M., Funke H., Reel J., Amoura Z., Pilette J.-C.,
 RA Delange J.-F., Brewer H.B., Duvenger N., Denelle P., Assmann G.,
 RT "Tangier disease is caused by mutations in the gene encoding
 RT ATP-binding cassette transporter 1".
 RL Nat. Genet. 22:352-355(1999).
 RN [17]
 RP VARIANTS FHA THR-1091 AND 1893-GLU-ASP-1894 DEL.
 RX MEDLINE=20001430; PubMed=10533863;
 RA Marcell M., Brooks-Wilson A., Clee S.M., Roomp K., Zhang L.-H., Yu L.,
 RA Collins J.A., van Dam M., Molhuizen H.O.F., Loubser O.,
 RA Ouellette B.F.F., Senses C.W., Fichter K., Ashbourne-Excoffon K.J.D.,
 RA Boucher B., Pimstone S., Genest J. Jr., Kastelein J.J.P., Hayden M.R.,
 RT "Mutations in the ABC1 gene in familial HDL deficiency with defective
 RT cholesterol efflux".
 RL Lancet 354:1341-1346(1999).
 RN [18]
 RP VARIANTS TD ARG-597 AND ARG-1477, AND VARIANT FHA LEU-693 DEL.
 RX MEDLINE=99364411; PubMed=10431236;
 RA Brooks-Wilson A., Marcell M., Clee S.M., Zhang L.-H., Roomp K.,
 RA van Dam M., Yu L., Brewer C., Collins J.A., Molhuizen H.O.F.,
 RA Loubser O., Ouellette B.F.F., Fichter K., Ashbourne-Excoffon K.J.D.,
 RA Senses C.W., Scherer S., Mott S., Denis M., Martindale D.,
 RA Frohlich J., Morgan K., Koop B., Pimstone S., Kastelein J.J.P.,
 RA Hayden M.R.,
 RT "Mutations in ABC1 in Tangier disease and familial high-density
 RT lipoprotein deficiency".
 RL Nat. Genet. 22:336-345(1999).
 RN [19]
 RP VARIANTS TD SER-590; SER-935 AND VAL-937, AND VARIANTS ALA-399 AND
 RX MET-883.
 RA MEDLINE=99364412; PubMed=10431237;
 RA Bodzioch M., Orso E., Klucken J., Langmann T., Boettcher A.,
 RA Diedrich W., Drobnik W., Barlage S., Buechler C.,
 RA Porcsch-Oezueremmez M., Kaminski W.E., Hehmann H.W., Oette K.,
 RA Rothe G., Aslanidis C., Lackner K.J., Schmitz G.,
 RT "The gene encoding ATP-binding cassette transporter 1 is mutated in
 RT Tangier disease".
 RL Nat. Genet. 22:347-351(1999).
 RN [10]
 RP VARIANTS TD ILE-929; ARG-597 AND ARG-1477, AND VARIANTS FHA LEU-693
 RX DEL; THR-1091; 1893-GLU-ASP-1894 DEL AND LEU-2150.
 RA MEDLINE=20540002; PubMed=11086027;
 RA Clee S.M., Kastelein J.J.P., van Dam M., Marcell M., Roomp K.,
 RA Zwarts K.Y., Collins J.A., Roelants R., Tamasa N., Stulic T.,
 RA Suda T., Ceska R., Boucher B., Rondeau C., Desoulh C.,
 RA Brooks-Wilson A., Molhuizen H.O.F., Frohlich J., Genest J. Jr.,
 RA Hayden M.R.,
 RT "Age and residual cholesterol efflux affect HDL cholesterol levels and
 RT coronary artery disease in ABCA1 heterozygotes".
 RL J. Clin. Invest. 106:1263-1270(2000).
 RN [11]
 RP VARIANTS TD ASN-1289 AND HIS-1800.
 RX MEDLINE=20171564; PubMed=10706591;
 RA Brousseau M.E., Schaefer E.J., Dupuis J., Eustace B.,
 RA Van Berdevegh P., Goldkamp A.L., Thurston L.M., Fitzgerald M.G.,
 RA Yasek-Mckenna D., O'Neill G., Eberhart G.P., Weiffenbach B.,
 RA Ordovas J.M., Freeman M.W., Brown R.H. Jr., Gu J.Z.,
 RT "Novel mutations in the gene encoding ATP-binding cassette 1 in four
 RT Tangier disease kindreds".
 RL J. Lipid Res. 41:433-441(2000).
 RN [12]
 RP VARIANT TD ASP-1046, VARIANT FHA CYS-230, AND VARIANTS LYS-219;
 RX ILE-825; MET-883 AND LYS-1587.
 RA MEDLINE=20396633; PubMed=10938021;
 RA Wang J., Burnett J.R., Near S., Young K., Ziman B., Hanley A.J.G.,
 RA Connelly P.W., Harris S.B., Hegele R.A.,
 RT "Common and rare ABCA1 variants affecting plasma HDL cholesterol".
 RL Arterioscler. Thromb. Vasc. Biol. 20:1983-1989(2000).
 RN [13]
 RP VARIANT TD TRP-587, AND VARIANT LEU-2168.
 RX MEDLINE=21157002; PubMed=11257260;
 RA Bertolini S., Plasciotta L., Seri M., Cusano R., Cantafora A.,
 RA Calabresi L., Franceschini G., Ravazzolo R., Calandra S.,
 RT "A point mutation in ABC1 gene in a patient with severe premature
 RT coronary heart disease and mild clinical phenotype of Tangier
 RT disease".
 RL Atherosclerosis 154:599-605(2001).
 RN [14]
 RP VARIANTS LYS-219; MET-883 AND ASP-1172.
 RX MEDLINE=21157003; PubMed=11257261;
 RA Brousseau M.E., Bodzioch M., Schaefer E.J., Goldkamp A.L., Kielar D.,
 RA Probst M., Ordovas J.M., Aslanidis C., Lackner K.J.,
 RA Bloomfield Robbins H., Collins D., Robins S.J., Wilson P.W.F.,
 RA Schmitz G.,
 RT "Common variants in the gene encoding ATP-binding cassette transporter
 RT 1 in men with low HDL cholesterol levels and coronary heart disease".
 RL Atherosclerosis 154:607-611(2001).
 RN [15]
 RP VARIANT TD LEU-1506.
 RX MEDLINE=21369429; PubMed=11476961;
 RA Lapicka-Bodzioch K., Bodzioch M., Kruehl M., Kielar D., Probst M.,
 RA Klec B., Andrikovics H., Boettcher A., Hubacek J., Aslanidis C.,
 RA Suttorp N., Schmitz G.,
 RT "Homogenous assay based on 52 primer sets to scan for mutations of
 RT the ABCA1 gene and its application in genetic analysis of a new
 RT patient with familial high-density lipoprotein deficiency syndrome".
 RL Biochim. Biophys. Acta 1537:42-48(2001).
 RN [16]
 RP VARIANTS TD ASN-1289 AND TRP-2081, AND VARIANT LYS-219.
 RX MEDLINE=21369433; PubMed=11476965;
 RA Huang W., Moriyama K., Koga T., Hua H., Ageta M., Kawabata S.,
 RA Mawatari K., Tamura F., Eto T., Kawamura M., Teramoto T., Sasaki J.,
 RT "Novel mutations in ABCA1 gene in Japanese patients with Tangier
 RT disease and familial high density lipoprotein deficiency with
 RT coronary heart disease".
 RL Biochim. Biophys. Acta 1537:71-78(2001).
 RN [17]
 RP VARIANTS LYS-219; ALA-399; MET-771; PRO-774; ASN-776; ILE-825;
 RX MET-883; ASP-1172; LYS-1587 AND CYS-1731.
 RA MEDLINE=21138379; PubMed=11238261;
 RA Clee S.M., Zwinderman A.H., Engert J.C., Zwarts K.Y.,
 RA Molhuizen H.O.F., Roomp K., Jukema J.W., van Miland M., van Dam M.,
 RA Hudson T.J., Brooks-Wilson A., Genest J. Jr., Kastelein J.J.P.,
 RA Hayden M.R.,
 RT "Common genetic variation in ABCA1 is associated with altered
 RT lipoprotein levels and a modified risk for coronary artery disease".
 RL Circulation 103:1198-1205(2001).
 RN [18]
 RP VARIANT TD THR-255, AND VARIANT ATHEROSCLEROSIS ASP-1611.
 RX MEDLINE=21645894; PubMed=11785958;
 RA Nishida Y., Hirano K., Tsukamoto K., Nagano M., Ikegami C., Roomp K.,
 RA Ishihara M., Sakane N., Zhang Z., Tsujii K., Matsuyama A., Ohama T.,
 RA Matsuura F., Ishigami M., Sakai N., Hiraoka H., Hattori H.,
 RA Wellington C., Yoshida Y., Misugi S., Hayden M.R., Egashira T.,
 RA Yamashita S., Matsuyama Y.,
 RT "Expression and functional analyses of novel mutations of ATP-binding
 RT cassette transporter-1 in Japanese patients with high-density
 RT lipoprotein deficiency".
 RL Biochem. Biophys. Res. Commun. 290:713-721(2002).
 CC -1- FUNCTION: CAMP-DEPENDENT AND SULFONYLUREA-SENSITIVE ANION
 CC TRANSPORTER. KEY GATEKEEPER INFLUENCING INTRACELLULAR CHOLESTEROL
 CC TRANSPORT.
 CC -1- TISSUE SPECIFICITY: WIDELY EXPRESSED, BUT MOST ABUNDANT IN
 CC MACROPHAGES.
 CC -1- DOMAIN: MULTIFUNCTIONAL POLYPEPTIDE WITH TWO HOMOLOGOUS HALVES,
 CC EACH CONTAINING AN HYDROPHOBIC MEMBRANE-ANCHORING DOMAIN AND AN
 CC ATP BINDING CASSETTE (ABC) DOMAIN.
 CC -1- DISEASE: DEFECTS IN ABCA1 ARE A CAUSE OF HIGH DENSITY LIPOPROTEIN
 CC DEFICIENCY TYPE I (HDL1), ALSO KNOWN AS TANGIER DISEASE (TD). TD
 CC IS A RECESSIVE DISORDER CHARACTERIZED BY ABSENCE OF HIGH DENSITY
 CC LIPOPROTEIN (HDL) CHOLESTEROL FROM PLASMA, HEPATOSPLENOMEGALY,
 CC PERIPHERAL NEUROPATHY, AND FREQUENTLY PREMATURE CORONARY ARTERY
 CC DISEASE (CAD).
 CC -1- DISEASE: Defects in ABCA1 are a cause of high density lipoprotein

Query Match 99.8%; Score 11770; DB 1; Length 2261;
 Best Local Similarity 99.7%; Pred. No. 0;
 Matches 2255; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 MACPOLRLMLKNNLFRFRRCQCLLEVAWPLFLLIISVRLSPYEQEHCFPKA 60
 Db 1 MACPOLRLMLKNNLFRFRRCQCLLEVAWPLFLLIISVRLSPYEQEHCFPKA 60
 QY 61 MPSAGTLPWOGIICNANNPCFRPTPGAPGVGNFNKSIYARLFSDARLLIXSQDT 120
 Db 61 MPSAGTLPWOGIICNANNPCFRPTPGAPGVGNFNKSIYARLFSDARLLIXSQDT 120
 QY 121 SMKDMRYLRTLOOIKKSSSNLKLODFLVNNEFFSGFLHNLSLPKSTYDKMLRAVILH 180
 Db 121 SMKDMRYLRTLOOIKKSSSNLKLODFLVNNEFFSGFLHNLSLPKSTYDKMLRAVILH 180
 QY 181 KVFLOGYOLHTSLCNGSKSEMIQLODOEYSELGLPKKLAABERVLRSMNDILKPTL 240
 Db 181 KVFLOGYOLHTSLCNGSKSEMIQLODOEYSELGLPKKLAABERVLRSMNDILKPTL 240
 QY 241 RLTNSTSPFSKELAEATKTLHSLGLAOLFSMBSMSMDROEVMFLTNVSSSSSTOI 300
 Db 241 RLTNSTSPFSKELAEATKTLHSLGLAOLFSMBSMSMDROEVMFLTNVSSSSSTOI 300
 QY 301 YQAVSRIVCGHPGGGLIKSLNMYEDNNYKALFGNGTEDEAETFYDNSTTPYCNLDM 360
 Db 301 YQAVSRIVCGHPGGGLIKSLNMYEDNNYKALFGNGTEDEAETFYDNSTTPYCNLDM 360
 QY 361 NDESSPLSRITWKALKPLLVGKILYTPPTPATROYMAVNTKPOLAVFHLEMEELS 420
 Db 361 NDESSPLSRITWKALKPLLVGKILYTPPTPATROYMAVNTKPOLAVFHLEMEELS 420
 QY 421 PKWTMERNOSOMDLYRMLDSDRDHFEQOOLDIDMTAODIYAFIAKHEDEVOSSNGS 480
 Db 421 PKWTMERNOSOMDLYRMLDSDRDHFEQOOLDIDMTAODIYAFIAKHEDEVOSSNGS 480
 QY 481 VYTWREAFNETNOAIRTISRMECVNLKLEPLATEVWLINXSMLEDBERKFMAGIVTGG 540
 Db 481 VYTWREAFNETNOAIRTISRMECVNLKLEPLATEVWLINXSMLEDBERKFMAGIVTGG 540
 QY 541 ITTGSIELPHHYKTKIRMDIDNVERTNKIKDGYWDPGRADPEEDMRYVWGFAVLODV 600
 Db 541 ITTGSIELPHHYKTKIRMDIDNVERTNKIKDGYWDPGRADPEEDMRYVWGFAVLODV 600
 QY 601 EOALIRVLGTGTEKKTGYMOQMPYPCYVDDIFLRVNSRSMPLFMTLAWITSAVAILKGI 660
 Db 601 EOALIRVLGTGTEKKTGYMOQMPYPCYVDDIFLRVNSRSMPLFMTLAWITSAVAILKGI 660
 QY 661 YEKARLKEHMRIMGLDNLISLFWISLIPLLVSAGLLVILKGNLLPYSDPSVFEV 720
 Db 661 YEKARLKEHMRIMGLDNLISLFWISLIPLLVSAGLLVILKGNLLPYSDPSVFEV 720
 QY 721 FLVSFAVVTILQCFILISTFSRANLAAAGGIIFTLRYLVCAMODYGVTLLKIFAS 780
 Db 721 FLVSFAVVTILQCFILISTFSRANLAAAGGIIFTLRYLVCAMODYGVTLLKIFAS 780
 QY 781 LLSVPAFGGCEYRALFEEOGIGVOMDNLFPSPVEEDGFNLITISISMULDFTLYGVM 840
 Db 781 LLSVPAFGGCEYRALFEEOGIGVOMDNLFPSPVEEDGFNLITISISMULDFTLYGVM 840
 QY 841 YLEAFPCOYGIIPRWYPCPKSYWFGESDEKSHPGSNOKRMEJOMEDEPHLKIGVS 900
 Db 841 YLEAFPCOYGIIPRWYPCPKSYWFGESDEKSHPGSNOKRMEJOMEDEPHLKIGVS 900
 QY 901 IONLVKVRDGMKAVAVDGLALNFYEGQITTSFLGHNAGKTTNSILGLPPTSGATYL 960
 Db 901 IONLVKVRDGMKAVAVDGLALNFYEGQITTSFLGHNAGKTTNSILGLPPTSGATYL 960
 QY 961 GKDIREKSTIRONLGVCPQHNVLFDMILVEEHIMFYARLKGLSEKHAKAMEOMALDVG 1020
 Db 961 GKDIREKSTIRONLGVCPQHNVLFDMILVEEHIMFYARLKGLSEKHAKAMEOMALDVG 1020

QY 1021 LPSSKLSKTSQLSGGMOKKLSVALAFVGGSVVILDEPTAGVDPYSRKGITWELLKYRQ 1080
 Db 1021 LPSSKLSKTSQLSGGMOKKLSVALAFVGGSVVILDEPTAGVDPYSRKGITWELLKYRQ 1080
 QY 1081 GRTIILSTHHMDEADVLDGRILAIISHKLCVGSGLFLKNOLGTGYTLTVKKNVESLS 1140
 Db 1081 GRTIILSTHHMDEADVLDGRILAIISHKLCVGSGLFLKNOLGTGYTLTVKKNVESLS 1140
 QY 1141 SCRNSSVSVYLLKEDSVQSSDAGLSDHESDPTLTDVSAISMLIRKHSEARLYEDI 1200
 Db 1141 SCRNSSVSVYLLKEDSVQSSDAGLSDHESDPTLTDVSAISMLIRKHSEARLYEDI 1200
 QY 1201 GHELTLYLPPEAKKEGAFVLFHEIDRLSDLGISYGISETTLEBFLKVAEEGVAE 1260
 Db 1201 GHELTLYLPPEAKKEGAFVLFHEIDRLSDLGISYGISETTLEBFLKVAEEGVAE 1260
 QY 1261 TSDGTLPARRRRAFGKQSCLEPFTEDDAADPNDSIDIPESRETDLLSGMKGCSYQK 1320
 Db 1261 TSDGTLPARRRRAFGKQSCLEPFTEDDAADPNDSIDIPESRETDLLSGMKGCSYQK 1320
 QY 1321 GMLTLOOQFVALLKRLILARRSKGFFAOIYLPVAFVCIALVESLIVPEGKYPSLEQ 1380
 Db 1321 GMLTLOOQFVALLKRLILARRSKGFFAOIYLPVAFVCIALVESLIVPEGKYPSLEQ 1380
 QY 1381 PMWNEQYTFVSNDAPEDTGTLLELNLTKDPGPGTCMGMPILPTPOAGEEEMTAP 1440
 Db 1381 PMWNEQYTFVSNDAPEDTGTLLELNLTKDPGPGTCMGMPILPTPOAGEEEMTAP 1440
 QY 1441 VQOTIMDLFONGNMTNOMPSPACOCSSDKIKMLPYCPGAGGLPPQRRONTADILQDL 1500
 Db 1441 VQOTIMDLFONGNMTNOMPSPACOCSSDKIKMLPYCPGAGGLPPQRRONTADILQDL 1500
 QY 1501 TGRNISDYLVKTYVOIILAKSLKNKIWNEFRYGGFSLGVSNTQALPPSOBYNAIKOMK 1560
 Db 1501 TGRNISDYLVKTYVOIILAKSLKNKIWNEFRYGGFSLGVSNTQALPPSOBYNAIKOMK 1560
 QY 1561 HLKLAODSSADRFNLISLGFMTGLDTRNNVYKWFENKGMHAISFLVYINNALILANLQK 1620
 Db 1561 HLKLAODSSADRFNLISLGFMTGLDTRNNVYKWFENKGMHAISFLVYINNALILANLQK 1620
 QY 1621 GENPSHYGTAENHPLNLTKQOLSEVALMTTSVDVLYSTICYFASFPASFFVFLQER 1680
 Db 1621 GENPSHYGTAENHPLNLTKQOLSEVALMTTSVDVLYSTICYFASFPASFFVFLQER 1680
 QY 1681 VSKAKHLQFISGKAPVYIMLSNFWDMCNVVPATLVIIIFCFOOKSVSSTNLPLVAL 1740
 Db 1681 VSKAKHLQFISGKAPVYIMLSNFWDMCNVVPATLVIIIFCFOOKSVSSTNLPLVAL 1740
 QY 1741 LLLLYGMSITPLMYPASFVFKIPSTAYVVLTSVNLFTIGINSVAFVLELFTDNKLNIN 1800
 Db 1741 LLLLYGMSITPLMYPASFVFKIPSTAYVVLTSVNLFTIGINSVAFVLELFTDNKLNIN 1800
 QY 1801 DILKSVLIPHPHCLGLIDMWKNOAMADALERPENFVSPLSMDLVRMLPMAVVG 1860
 Db 1801 DILKSVLIPHPHCLGLIDMWKNOAMADALERPENFVSPLSMDLVRMLPMAVVG 1860
 QY 1861 VVFFLITVLYOYRFFIPRPVNAKLSPLNDEDEDVRRERORIIDGGGONDILEIKELTKI 1920
 Db 1861 VVFFLITVLYOYRFFIPRPVNAKLSPLNDEDEDVRRERORIIDGGGONDILEIKELTKI 1920
 QY 1921 YRRKRPVADRICVGIIPPEGCGLLGVNAGKSTFEMLIGDITTVIRGDAFLKNKSILSN 1980
 Db 1921 YRRKRPVADRICVGIIPPEGCGLLGVNAGKSTFEMLIGDITTVIRGDAFLKNKSILSN 1980
 QY 1981 THEVONNNGYCOFPAITELLTGREHVEFFALLRGVBEKEVGKVEWAIIRKLCVYKGEK 2040
 Db 1981 THEVONNNGYCOFPAITELLTGREHVEFFALLRGVBEKEVGKVEWAIIRKLCVYKGEK 2040
 QY 2041 YAGNYSOGKRRKRLSTAMALLGGPPVYVFLDEPTTGMDKARREFLWNCALSVYKBEGRSVLT 2100
 Db 2041 YAGNYSOGKRRKRLSTAMALLGGPPVYVFLDEPTTGMDKARREFLWNCALSVYKBEGRSVLT 2100
 QY 2101 SHSMECEALCIRMAIMVNGRFRCLGSVQHLKNRFGDGYTIVVRIAGSNPDLKPVQDFPG 2160

Db 2101 SHMECEALCTMAIMVNGRERCLSGVOHKNRFGDGYTVIRIAGSNPDLKPYDQFEG 2160
 QY 2161 LAFPGSVLEKHEHNMLOVLPSSLARIFSLISOSKRLHIEYVSOTLDOVYNF 2220
 Db 2161 LAFPGSVPEKHEHNMLOVLPSSLARIFSLISOSKRLHIEYVSOTLDOVYNF 2220
 QY 2221 AKQSDDDLKDLSTLHKNOTVDVAVLSFLQDEKVESYV 2261
 Db 2221 AKQSDDDLKDLSTLHKNOTVDVAVLSFLQDEKVESYV 2261

RESULT 2

ACBL_MOUSE

ID ACBL_MOUSE STANDARD: PRT: 2261 AA.

AC P41233;

DT 01-FEB-1995 (Rel. 31, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE ATP-binding cassette, sub-family A, member 1 (ATP-binding cassette transporter 1) (ATP-binding cassette 1) (ABC-1).

GN ABCA1 OR ABC1.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=DA/2; TISSUE=Macrophage;

RA MEDLINE=94375008; PubMed=8088782;

RT Luciani M.F., Denizot F., Savary S., Mattei M.-G., Chimini G.;

RT "Cloning of two novel ABC transporters mapping on human chromosome 9.";

RL Genomics 21:150-159(1994).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J;

RA Qiu Y., Cavelier L., Chiu S., Rubin E., Cheng J.-F.;

RT "Human and mouse ABCA1 comparative sequencing and transgenesis studies identify potential regulatory sequences.";

RL Submitted (Jul-2000) to the EMBL/Genbank/DBJ databases.

CC - FUNCTION: CAMP-DEPENDENT AND SULFONYLUREA-SENSITIVE ANION TRANSPORTER. KEY GATEKEEPER INFLUENCING INTRACELLULAR CHOLESTEROL TRANSPORT (BY SIMILARITY).

CC - TISSUE SPECIFICITY: WIDELY EXPRESSED IN ADULT TISSUES. HIGHEST LEVELS ARE FOUND IN PREGNANT UTERUS AND UTERUS.

CC - DOMAIN: MULTIFUNCTIONAL POLYPEPTIDE WITH TWO HOMOLOGOUS HALVES, EACH CONTAINING AN HYDROPHOBIC MEMBRANE-ANCHORING DOMAIN AND AN -I- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. ABCA SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial or send an email to license@isb-sib.ch).

EMBL: X75926; CAAS3530.1; ALT_INIT.
 EMBL: AF287263; AAC39073.1; ALT_INIT.
 MGD: MGI:99607; Abcal.
 InterPro: IPR003593; AAA_ATPase.
 InterPro: IPR003439; ABC_transporter.
 Pfam: PF00005; ABC_tran; 2.
 Prodom: PD000006; ABC_transporter; 2.
 SMART: SM00382; AAA; 1.
 PROSITE: PS00211; ABC_TRANSPORTER; 1.
 KATP-binding; Glycoprotein; Transmembrane; Transport.

FT TRANSMEM 749 765
 FT TRANSMEM 771 787
 FT TRANSMEM 1041 1057
 FT TRANSMEM 1351 1367
 FT TRANSMEM 1661 1677
 FT TRANSMEM 1708 1724
 FT TRANSMEM 1737 1753
 FT TRANSMEM 1775 1791
 FT TRANSMEM 1854 1870
 FT NP_BIND 933 940
 FT NP_BIND 1946 1953
 FT CARBOHYD 14 14
 FT CARBOHYD 98 98
 FT CARBOHYD 151 151
 FT CARBOHYD 161 161
 FT CARBOHYD 196 196
 FT CARBOHYD 244 244
 FT CARBOHYD 292 292
 FT CARBOHYD 337 337
 FT CARBOHYD 349 349
 FT CARBOHYD 400 400
 FT CARBOHYD 478 478
 FT CARBOHYD 489 489
 FT CARBOHYD 521 521
 FT CARBOHYD 820 820
 FT CARBOHYD 1144 1144
 FT CARBOHYD 1294 1294
 FT CARBOHYD 1453 1453
 FT CARBOHYD 1499 1499
 FT CARBOHYD 1504 1504
 FT CARBOHYD 1637 1637
 FT CARBOHYD 2044 2044
 FT CARBOHYD 2238 2238
 FT CONFLICT 1567 1568
 FT CONFLICT 2024 2024
 SQ SEQUENCE 2261 AA; 254011 MW; FA62B21FD1D09F9 CRC64;

Query Match
 Best Local Similarity 95.4%; Score 11256; DB 1; Length 2261;
 Matches 2149; Conservative 55; Mismatches 57; Indels 0; Gaps 0;

QY 1 MACMPOQLRLMLKNTFRROTCOLLLEVAMPFLFILISVRLSPYEDHCHPKNKA 60
 Db 1 MACMPOQLRLMLKNTFRROTCOLLLEVAMPFLFILISVRLSPYEDHCHPKNKA 60
 QY 61 MPSAGTLPWVGIIICMANNPCFRYPFGAEPGVGNKSIYRLSPDARLLYSQKT 120
 Db 61 MPSAGTLPWVGIIICMANNPCFRYPFGAEPGVGNKSIYRLSPDARLLYSQKT 120
 QY 121 SMKDMRKVLRTLOQIRKSSNLIKLOPLVDNEFSGFLYHNLSLPKSTVDKMLRADVILH 180
 Db 121 SMKDMRKVLRTLOQIRKSSNLIKLOPLVDNEFSGFLYHNLSLPKSTVDKMLRADVILH 180
 QY 181 KVFLOGLHLLTSLCNGSKSEEMIOGLDOEVSEFLGCPREKILAAAEVRLSRMDLKITL 240
 Db 181 KVFLOGLHLLTSLCNGSKSEEMIOGLDOEVSEFLGCPREKILAAAEVRLSRMDLKITL 240
 QY 241 RLINSTSPFSKELAEVTLTLHSLGTLAQELFSRMSXDMROEVPLTNSSSSSTOI 300
 Db 241 RLINSTSPFSKELAEVTLTLHSLGTLAQELFSRMSXDMROEVPLTNSSSSSTOI 300
 QY 301 YQAVSRIVCGHPGGLKIKSLNWEEDNNYKALFGGNGCEDEAFEPYNSNTPPCNDLMK 360
 Db 301 YQAVSRIVCGHPGGLKIKSLNWEEDNNYKALFGGNGCEDEAFEPYNSNTPPCNDLMK 360
 QY 361 NLESSLSTIIKALKPLVGLKILTPPTPATROYMAEVNKTFOELAVFHDLEBMEELS 420
 Db 361 NLESSLSTIIKALKPLVGLKILTPPTPATROYMAEVNKTFOELAVFHDLEBMEELS 420
 QY 421 PKIWTFMNSQEMDLVRMLDLSRDNDHFWEQOLGDMFTADDIYAFILAKHPEDVOSSNGS 480
 Db 421 PKIWTFMNSQEMDLVRMLDLSRDNDHFWEQOLGDMFTADDIYAFILAKHPEDVOSSNGS 480

QY 481 VYTWREAFNETQAIRTISRMECYNNKLBPATEVWLKNSMELDERKFMAGIVFTG 540
 DB 481 VYTWREAFNETQAIOTISRMECYNNKLBPATEVWLKNSMELDERKFMAGIVFTG 540
 QY 541 IFPGSIELPHHYKIRMDIDVERTNKIKDGYWDPGPADPFEDMRYWGGFAYLDQV 600
 DB 541 ITDSVELPHHYKIRMDIDVERTNKIKDGYWDPGPADPFEDMRYWGGFAYLDQV 600
 QY 601 EOAIRVLGTEKKTGYVMOQPRPCYVDIFLRVSRSMPLFMTLANIYSAVATIKGIV 660
 DB 601 EOAIRVLGTEKKTGYVMOQPRPCYVDIFLRVSRSMPLFMTLANIYSAVATIKGIV 660
 QY 661 YEKARLKETMRIMGDNGILMFSWVSSILPLVASAGLIVILKGNLPLSDSVYFV 720
 DB 661 YEKARLKETMRIMGDNGILMFSWVSSILPLVASAGLIVILKGNLPLSDSVYFV 720
 QY 721 FLVSFAVVTILQCFILISTLFSRANLAAAGGITYFTLYLPYLVCMNOGYVFTLIKIFAS 780
 DB 721 FLVSFAVVTILQCFILISTLFSRANLAAAGGITYFTLYLPYLVCMNOGYVFTLIKIFAS 780
 QY 781 LLSVAFGFGCEYFALFEEGIGVOMDNLFESPVEEDGFNLTTISMLFDTFLYGVMTW 840
 DB 781 LLSVAFGFGCEYFALFEEGIGVOMDNLFESPVEEDGFNLTTISMLFDTFLYGVMTW 840
 QY 841 YIEAVFPGQYGIIPRPWYFPCTKSYWGESEDEKSHPGSNOKRMSLCEMEEPTHLKIGVS 900
 DB 841 YIEAVFPGQYGIIPRPWYFPCTKSYWGESEDEKSHPGSNOKRMSLCEMEEPTHLKIGVS 900
 QY 901 IONLVKVRDGMKVAVDSALNFEGOITSLFNGAGKTTMSIITGLPPTSGTAYIL 960
 DB 901 IONLVKVRDGMKVAVDSALNFEGOITSLFNGAGKTTMSIITGLPPTSGTAYIL 960
 QY 961 GKDISEMSTIRONLGVCPHNVLFDMLTVEHIMFYARLKIGSEKHYVAKEMEDMALDVG 1020
 DB 961 GKDISEMSTIRONLGVCPHNVLFDMLTVEHIMFYARLKIGSEKHYVAKEMEDMALDVG 1020
 QY 1021 LPSSKLKSTQSLSGMOKRLSVLAFYVGSVYLDEPTAGVPRSGRGITWELLKYRQ 1080
 DB 1021 LPSSKLKSTQSLSGMOKRLSVLAFYVGSVYLDEPTAGVPRSGRGITWELLKYRQ 1080
 QY 1081 GRTIILSTHMHMEDADYLGRILAIISHGKLCVGSLLFKNLGTYGTYLTLVKKVESLS 1140
 DB 1081 GRTIILSTHMHMEDADYLGRILAIISHGKLCVGSLLFKNLGTYGTYLTLVKKVESLS 1140
 QY 1141 SCRNSSTVSYLKEDSVSSDAGLSDHSDPLTIDVSAISLIRKHVSEARLYEDI 1200
 DB 1141 SCRNSSTVSYLKEDSVSSDAGLSDHSDPLTIDVSAISLIRKHVSEARLYEDI 1200
 QY 1201 GHLEITVLPPEAKKEGAFELFHEIDRLSDIGISSYGETTLEIFLKAEEGVDAE 1260
 DB 1201 GHLEITVLPPEAKKEGAFELFHEIDRLSDIGISSYGETTLEIFLKAEEGVDAE 1260
 QY 1261 TSDGTLPARRNRRAFGDKOSCLPFTEDDAADPNDSIDPESRENDILSGMKGSYOVK 1320
 DB 1261 TSDGTLPARRNRRAFGDKOSCLPFTEDDAADPNDSIDPESRENDILSGMKGSYOVK 1320
 QY 1321 GMRKLTQOOFVALLMKRLIRSRKGFPAQIYPAVFCIALVFSILVPPFGKPSLEIQ 1380
 DB 1321 GMRKLTQOOFVALLMKRLIRSRKGFPAQIYPAVFCIALVFSILVPPFGKPSLEIQ 1380
 QY 1381 PMWNEQYTVSVNDAPEDTGTLEILNALIKDPCGFTGRCHMGNIPTPTQOAGEEEMTAP 1440
 DB 1381 PMWNEQYTVSVNDAPEDTGTLEILNALIKDPCGFTGRCHMGNIPTPTQOAGEEEMTAP 1440
 QY 1441 VPQITMDLFONGMNTMNPSPACOCSSDKIKKMLPVCPPAGGILPPOROKOTADILQDL 1500
 DB 1441 VPQITMDLFONGMNTMNPSPACOCSSDKIKKMLPVCPPAGGILPPOROKOTADILQDL 1500
 QY 1501 TGRNISDLVKTYYOQIILAKSLKNKIWNEFRYGGFSLGVSNTQALPPSOEVNDAIKOMKK 1560
 DB 1501 TGRNISDLVKTYYOQIILAKSLKNKIWNEFRYGGFSLGVSNTQALPPSOEVNDAIKOMKK 1560
 QY 1561 HLKLAKSSADRLNLSLGRFMTGLDTRNNYVWFMNKGWHAISSEFLVNNAILRANLQK 1620

DB 1561 LKLTQTSADFLFSLGRFMAGLDTKNNVWFMNKGWHAISSEFLVNNAILRANLQK 1620
 QY 1621 GENPSHYGITAENHPLNLTQOOLSEVALMTTSVDVLYSICVIFAMSPVPASFVFLQER 1680
 DB 1621 GENPSHYGITAENHPLNLTQOOLSEVALMTTSVDVLYSICVIFAMSPVPASFVFLQER 1680
 QY 1681 VSKAKHLOFISGVKPYIYLSNFWDMCNVYVPAVLVIIIFICFOQKSYVSTNLPVAL 1740
 DB 1681 VSKAKHLOFISGVKPYIYLSNFWDMCNVYVPAVLVIIIFICFOQKSYVSTNLPVAL 1740
 QY 1741 LLLLYGWSITPLMPASFVEFKIPSTAYVLTISVNLFIGINGSVATFVLEIFTPDKINNIN 1800
 DB 1741 LLLLYGWSITPLMPASFVEFKIPSTAYVLTISVNLFIGINGSVATFVLEIFTPDKINNIN 1800
 QY 1801 DILKSVLIFPHHCGLRGILDMYKNOAMADALERGEENFVSPLSMDLVGRNLFMAAVEG 1860
 DB 1801 DILKSVLIFPHHCGLRGILDMYKNOAMADALERGEENFVSPLSMDLVGRNLFMAAVEG 1860
 QY 1861 VVFFLITVLIOYRFFIRPREVNAKLSPUNDEDEDVREBQRILLDGGQNDILEIKELTKI 1920
 DB 1861 VVFFLITVLIOYRFFIRPREVNAKLSPUNDEDEDVREBQRILLDGGQNDILEIKELTKI 1920
 QY 1921 YRRKRPAVDRIKICGIPGCEFGILGVNAGKSTKMLTGTPTVYRGDAFLNKSILSN 1980
 DB 1921 YRRKRPAVDRIKICGIPGCEFGILGVNAGKSTKMLTGTPTVYRGDAFLNKSILSN 1980
 QY 1981 IHEHONNGYCPQDAILTELTGREHEVFPALLRGVEKEVKGEMAIRKGLYKYGK 2040
 DB 1981 IHEHONNGYCPQDAILTELTGREHEVFPALLRGVEKEVKGEMAIRKGLYKYGK 2040
 QY 2041 YAGNYSNGNRKRLSTANALJGGPPVYFLDEPTGMDPKARFLMNCALSYVKGESVLT 2100
 DB 2041 YAGNYSNGNRKRLSTANALJGGPPVYFLDEPTGMDPKARFLMNCALSYVKGESVLT 2100
 QY 2101 SHMECEALCTRNAIMVNGFRCLGSVOHLKNGFGGYTIVARIAGSNPDLKPVODFFG 2160
 DB 2101 SHMECEALCTRNAIMVNGFRCLGSVOHLKNGFGGYTIVARIAGSNPDLKPVODFFG 2160
 QY 2161 LAEPGSVLKEKRNMLQYOLPSSLSLARITSIISQSKRRLHIEDYSVOTTLDOYVNF 2220
 DB 2161 LAEPGSVLKEKRNMLQYOLPSSLSLARITSIISQSKRRLHIEDYSVOTTLDOYVNF 2220
 QY 2221 AKDQSDHKLKDLSLKKNQTVDAVLTSLFLODEKVESY 2261
 DB 2221 AKDQSDHKLKDLSLKKNQTVDAVLTSLFLODEKVESY 2261

RESULT 3
 ABR_C_HUMAN STANDARD: PRT: 2273 AA.
 ID ABR_C_HUMAN 578363; 060915; 015112;
 AC 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Retinal-specific ATP-binding cassette transporter (RIM ABC
 DE transporter) (RIM protein) (RMP) (Stargardt disease protein).
 GN ABCA4 OR ABR4.
 OS Homo sapiens (human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 OC NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A., VARIANTS STGD, AND VARIANTS HIS-846 AND GLN-943.
 RX MEDLINE-97207641; PubMed-9054934;
 RA Allikmets R., Singh N., Sun H., Shroyer N.F., Hutchinson A.,
 RA Chidambaram A., Getraud B., Baird L., Stauffer D., Pellicer A.,
 RA Rattiner A., Smallwood P.M., Li Y., Anderson K.L., Lewis R.A.,
 RA Nathans J., Leppert M., Dean M., Lupski J.R.,
 RA "A photoreceptor cell-specific ATP-binding transporter gene (ABCA4) is
 RT mutated in recessive Stargardt macular dystrophy."
 RL Nat. Genet. 15:236-246(1997).
 RN [2]

- RP SEQUENCE FROM N.A.
RX MEDLINE=97345663; PubMed=9202155;
RA Azarian S.M., Travis G.H.;
RT "The photoreceptor rim protein is an ABC transporter encoded by the
RL gene for recessive Stargardt's disease (ABCR).";
RN FEBS Lett. 409:247-252(1997).
[3]
RX SEQUENCE FROM N.A., AND VARIANTS STGD TRP-18 AND CYS-212.
RX MEDLINE=98163759; PubMed=9503029;
RA Gerber S., Rozet J.-M., van de Pol T.J.R., Hoyng C.B., Munnich A.,
RA Blankenagel A., Kaplan J., Cremers F.P.M.;
RT "Complete exon-intron structure of the retina-specific ATP binding
RT transporter gene (ABCR) allows the identification of novel mutations
RL underlying Stargardt disease.";
RN Genomics 48:139-142(1998).
[4]
RX SEQUENCE FROM N.A., AND VARIANTS STGD.
RX MEDLINE=98141123; PubMed=9490294;
RA Nasonkin I., Tilling M., Koehler M.R., Schmidt M., Molday R.S.,
RA Weber B.H.F.;
RT "Mapping of the rod photoreceptor ABC transporter (ABCR) to 1p21-p22.1
RT and identification of novel mutations in Stargardt's disease.";
RN Hum. Genet. 102:21-26(1998).
[5]
RX CHARACTERIZATION.
RX MEDLINE=99175213; PubMed=10075733;
RA Sun H., Molday R.S., Nathans J.;
RT "Retinal stimulates ATP hydrolysis by purified and reconstituted ABCR,
RT the photoreceptor-specific ATP-binding cassette transporter
RL responsible for Stargardt disease.";
RN J. Biol. Chem. 274:8269-8281(1999).
[6]
RX DISEASE.
RX MEDLINE=98133912; PubMed=9466990;
RA Cremers F.P.M., van de Pol D.J.R., van Driel M.A., den Hollander A.I.,
RA van Haren F.J.J., Koeber N.V.A.M., Tijmes N., Bergen A.A.B.,
RA Rohrschneider K., Blankenagel A., Plinckers A.O.L.G., Deutman A.F.,
RA Hoyng C.B.;
RT "Autosomal recessive retinitis pigmentosa and cone-rod dystrophy
RT caused by splice site mutations in the Stargardt's disease gene
RL ABCR.";
RN Hum. Mol. Genet. 7:355-362(1998).
[7]
RX VARIANTS ARM2D, AND VARIANTS.
RX MEDLINE=97442530; PubMed=9295268;
RA Allikmeets R., Shroyer N.F., Singh N., Seddon J.M., Lewis R.A.,
RA Bernstein P.S., Peiffer A., Zabitskie N.A., Li Y., Hutchinson A.,
RA Dean M., Lupski J.R., Leppert M.;
RT "Mutation of the Stargardt disease gene (ABCR) in age-related macular
RT degeneration.";
RN Science 277:1805-1807(1997).
[8]
RX VARIANTS STGD W-18; C-212; H-636; M-1019; V-1038; C-1108; W-1640;
RX S-1977 AND H-2107, AND VARIANTS FEM P-11; P-541; V-1038; E-1091;
RP C-1508; F-1970 AND R-1971
RX MEDLINE=98454319; PubMed=9781034;
RA Rozet J.-M., Gerber S., Souied E., Perrault I., Chatelain S., Ghazi I.,
RA Leewski C., Dullier J.-L., Munnich A., Kaplan J.;
RT "Spectrum of ABCR gene mutations in autosomal recessive macular
RT dystrophies.";
RN Eur. J. Hum. Genet. 6:291-295(1998).
[9]
RX VARIANTS STGD.
RX MEDLINE=99138655; PubMed=9973280;
RA Lewis R.A., Shroyer N.F., Singh N., Allikmeets R., Hutchinson A.,
RA Li Y., Lupski J.R., Leppert M., Dean M.;
RT "Genotype/phenotype analysis of a photoreceptor-specific ATP-binding
RT cassette transporter gene, ABCR, in Stargardt disease.";
RN Am. J. Hum. Genet. 64:422-434(1999).
[10]
RX VARIANTS STGD, AND VARIANTS.
RX MEDLINE=99192348; PubMed=10090887;
Maugeri A., van Driel M.A., van de Pol D.J.R., Klievering B.J.,
RP van Haren F.J.J., Tijmes N., Bergen A.A.B., Rohrschneider K.,
RA Blankenagel A., Plinckers A.J.L.G., Dahl N., Brunner H.G.,
RA Deutman A.F., Hoyng C.B., Cremers F.P.M.;
RT "The 2586G->C mutation in the ABCR gene is a mild frequent founder
RT mutation in the western European population and allows the
RT classification of ABCR mutations in patients with Stargardt disease.";
RN Am. J. Hum. Genet. 64:1024-1035(1999).
[11]
RX VARIANTS STGD TYR-54, AND VARIANT ALA-863.
RX MEDLINE=20077755; PubMed=10612508;
RA Zhang K., Garibaldi D.C., Kuzava M., Albini T., Chiang M.F.,
RA Kerrigan M., Sunness J.S., Han M., Allikmeets R.;
RT "A novel mutation in the ABCR gene in four patients with autosomal
RT recessive Stargardt disease.";
RN Am. J. Ophthalmol. 128:720-724(1999).
[12]
RX VARIANTS STGD V-60; R-206; N-300; P-541; A-849; P-974; V-1038; C-1108;
RP L-1408; R-1488; D-1652; P-1729; E-1961; W-2038; W-2077; H-2107; R-2128
RP AND Y-2150.
RX MEDLINE=99221420; PubMed=10206579;
RA Fishman G.A., Stone E.M., Grover S., Derlacki D.J., Haines H.L.,
RA Hockey R.R.;
RT "Variation of clinical expression in patients with Stargardt dystrophy
RT and sequence variations in the ABCR gene.";
RN Arch. Ophthalmol. 117:504-510(1999).
[13]
RX VARIANTS GLU-1961 AND ASN-2177.
RX MEDLINE=20349288; PubMed=10880298;
RA Allikmeets R., Tammur J., Hutchinson A., Lewis R.A., Shroyer N.F.,
RA Dalakishvili K., Lupski J.R., Steiner K., Paulschoff D., Holz F.G.,
RA Singh N., Peiffer A., Zabitskie N.A., Leppert M., Seddon J.M.,
RA Zhang K., Sunness J.S., Udar N.S., Yelchits S., Silva-Garcia R.,
RA Small K.W., Simonelli F., Testa F., D'Urso M., Brancato R.,
RA Rinaldi E., Iyegst S., Taube A., Wadelius C., Souied E., Ducrocq D.,
RA Kaplan J., Assink J.J.M., ten Brink J.B., de Jong P.T.V.M.,
RA Bergen A.A.B., Maugeri A., van Driel M.A., Hoyng C.B., Cremers F.P.M.,
RA Palma E., Coco R., Balcells S., Gonzalez-Duarte R., Kermani S.,
RA Stanga P., Bhattacharya S.S., Bird A.C.;
RT "Further evidence for an association of ABCR alleles with age-related
RT macular degeneration.";
RN Am. J. Hum. Genet. 67:487-491(2000).
[14]
RX VARIANTS STGD E-60; T-60; E-65; L-68; R-72; C-212; S-230; S-247;
RP V-328; K-471; P-541; Q-572; R-607; K-635; C-653; Y-764; R-765; A-901;
RP I-959; K-1036; V-1038; P-1063; D-1087; C-1097; C-1108; L-1380; K-1399;
RP T-1430; V-1440; H-1443; L-1486; Y-1488; M-1537; P-1689; L-1705;
RP T-1733; R-1748; P-1763; K-1885; H-1898; E-1961; S-1977; G-2077
RP W-2077 AND V-2241, AND VARIANTS Q-152; H-212; R-423; I-552; R-914;
RP Q-943; T-1562; I-1868; M-1921; L-1948; F-1970; A-2059; N-2177 AND
RP V-2216.
RX MEDLINE=20442027; PubMed=10958763;
RA Rivera A., White K., Stoeck H., Steiner K., Hemmrich N., Grimm T.,
RA Jurkles B., Lorenz B., Scholl H.P.N., Apfelstedt-Sylla E.,
RA Weber B.H.F.;
RT "A comprehensive survey of sequence variation in the ABCA4 (ABCR) gene
RT in Stargardt disease and age-related macular degeneration.";
RN Am. J. Hum. Genet. 67:800-813(2000).
[15]
RX VARIANTS CORD3 GLU-65; CYS-212; PRO-541; ALA-863; GLY-863 DEL;
RP VAL-1038; LYS-1122; TYR-1490 AND ASP-1598.
RX MEDLINE=20442040; PubMed=10958761;
RA Maugeri A., Klievering B.J., Rohrschneider K., Blankenagel A.,
RA Brunner H.G., Deutman A.F., Hoyng C.B., Cremers F.P.M.;
RT "Mutations in the ABCA4 (ABCR) gene are the major cause of autosomal
RT recessive cone-rod dystrophy.";
RN Am. J. Hum. Genet. 67:960-966(2000).
[16]
RX VARIANTS STGD ASP-340; GLN-572; ALA-863; SER-965; VAL-1038; ALA-1780
RP AND HIS-1898, AND VARIANT GLN-943.
RX MEDLINE=20208356; PubMed=10746567;
RA Shroyer N.F., Lewis R.A., Lupski J.R.;
RT "Complex inheritance of ABCR mutations in Stargardt disease: linkage

RT disequilibrium, complex alleles, and pseudodominance";
 RL Hum. Genet. 106:244-248(2000).
 RN [17]
 RX VARIANTS STGD
 RX MEDLINE=20098082; PubMed=10634594;
 RA Bataillon M., Ocaña L., Bessant D., Lois N., Bird A.C., Payne A.,
 RA Bhatnagar S.S.;
 RT "An analysis of ABCR mutations in British patients with recessive
 RT retinal dystrophies";
 RL Invest. Ophthalmol. Vis. Sci. 41:16-19(2000).
 RN [18]
 RX VARIANTS STGD C-212; D-767; T-897; V-1038; K-1087; K-1399; Q-1640 AND
 RX E-1961, AND VARIANT HIS-212.
 RX MEDLINE=20174852; PubMed=10711710;
 RA Simionelli F., Testa F., de Crecholo G., Rinaldi E., Hutchinson A.,
 RA Atkinson A., Dean M., D'urso M., Allikmets R.;
 RT "New ABCR mutations and clinical phenotype in Italian patients with
 RT Stargardt disease";
 RL Invest. Ophthalmol. Vis. Sci. 41:892-897(2000).
 RN [19]
 RX CHARACTERIZATION OF VARIANTS, AND MUTAGENESIS OF GLY-966; LYS-969;
 RX GLY-1975 AND LYS-1978.
 RX MEDLINE=20472331; PubMed=11017087;
 RA Sun H., Smallwood P.M., Nathans J.;
 RT "Biochemical defects in ABCR protein variants associated with human
 RT retinopathies";
 RL Nat. Genet. 26:242-246(2000).
 RN [20]
 RX VARIANT STGD ASN-972, AND VARIANTS GLN-943; ILE-1868 AND LEU-1948.
 RX MEDLINE=21478761; PubMed=1159493;
 RA Eksanah L., Ekstrom U., Abrahamson M., Bauer B., Andreasson S.;
 Query Match 49.8%; Score 5870.5; DB 1; Length 2273;
 Best Local Similarity 50.0%; Pred. No. 0;
 Matches 1165; Conservative 358; Mismatches 635; Indels 171; Gaps 27;

DB 522 LDKFESYNDENQULQORALSLLEENMFAGVYFPMYFWTSSLPBHYKXKIRMDIDVERT 581
 567 NKIKDGYWDPGRADPEEDMRYWNGFAYLQDVYEOALIRVLTGTEKKRGVYQOMPYP 626
 DB 582 NKIKRWDYSGFRADPEEDFRIYWGFAVYLODVEOQITRSQVQAEAPVGIYLOQMPYPC 641
 627 YVDDIFLRYMSRMPLEFMTLAMIYSAVITIKGIYEEKARLEKEMRIMGDNSTLWFSW 666
 DB 642 EYDSEFMILNCFPLFMYLAMIYSAVITIKGIYEEKARLEKEMRIMGDNSTLWFSW 701
 687 ISSLPLVYAGLIVYILKGNLDPYSDPSVEYFVLSFAVAVITLQCFILISTLFSRANLA 746
 DB 702 LDFSEISMSTIFLTFIMHGRILHYSDPFLFLFLAFTATIMLCFLSTFEFSKASLA 761
 747 AAGCGIITVTLVPLVLCVAMQDVYGFLLKIFASLSLVAAGFCEYFALPEEGIGVOM 806
 DB 762 AACGVITFPLPHILCFAMQDMRTAEKRAVSLSLVAAGFCEYFALPEEGIGVOM 821
 807 DNLFSPEVERDGNLTSTSMILFDTFLYGVMTWYIEAVPFGOYGIIPRYFPCTKSTWF 866
 DB 822 SNIGNSPTGEDDEFSLMQMMLLAAYVGLAWYLDQVFPDQYGPPLPWYFLQESYWL 881
 867 G-----ESDEKSHPGSNOKRMSEICMEEPETHLKGVSIONLVKY 907
 DB 882 GCGGSTRERALEKTEPLETEDEPDPH---EGIHDSFFERHPGWPGVCYANLVKI 937
 908 YROGMKAVVAGLNLTFYEGQITSLFGNCGAKTTMTSILTGFLPPTSGTAYILGDNIRSE 967
 DB 938 FEPGGRPAVRLNITTYENDITFELHNGAGKTTLSITLGLPPTSGVALVAGRDIEFS 997
 968 MSTIRONLGVCPQHNVLFDMLVVEEHIWYARIKGLSEKHYVAEBQNALDVGLPSSKLK 1027
 DB 998 LDAVQSLGMCPOHNLIFHLLITVAEHMLFYAQLKGKSOEADLEWEMLEDTGL-HHNRN 1056
 1028 SKTSQLSGGMQKLSVALAVYGGSKVYIIDEPRYAGDPYSRSGIEMELLYKRGQRTIIS 1087
 DB 1057 EEAODLSGGMRKLSVALAVYGGSKVYIIDEPRYAGDPYSRSGIEMELLYKRGQRTIIS 1116
 1088 THHNDADVLADRIATLISHGLCCVSSFLKNQLOCTGYVTLVYKDVSSLSGCRNSS 1147
 DB 1117 THHNDADVLADRIATLISHGLCCVSSFLKNQLOCTGYVTLVYKDVSSLSGCRNSS 1175
 1148 TVSYLKEDSVSSSSDAGLCSHESDTLTIDVSALSNLRKVSARLYVEIGHETLV 1207
 DB 1176 TCSGSSKGFSTTCPAHVDDLTPEQVLDGVDNEIMDVVLAHVPKALVEGIGDELLFL 1232
 1208 LPEFAKGCAGVLEFHEIDRLSDLGSSYGISSETTLEELFLKVAEBSGVDAETSDGTL 1267
 DB 1233 LPNKFKAHAYASLFRLEETLADLGISFCISDTPLEELFLKVAEBSGVDAETSDGTL 1289
 1268 ARNRRAFQKQSCLRPTEDDAADPNDSDI-----DPSRETDLSCM 1311
 DB 1290 AOKRENNVNRHPCIGP-REKAGOTPODSNVCSGAPAAHREGOPPEPCPGPQINTGT 1348
 1312 DKGSGYQKMKLITQOQVVALNLRLLIARRSRKGFPAIVLPAVFCIALVSLIVPFF 1371
 DB 1349 -----OLVQIHVOLLVNRFOHTIRSHKDFLAQIVLPATVFLALMISYIAP 1397
 1372 GYPSLELOPMWYNEQYTFVSNDAPEDTGTELLNALTKDPGEGTCROMESNP1PDPPCA 1431
 DB 1398 GREPALITLHPWYIGQYTFEFSMDPEGSQFTVLADVLNLRKGFENRCLKSGMREYPC-G 1456
 1432 GDEEWTPAVPOTIMDLFQNGMTQONSPPACOCSSQKIKKMLPVPCCGAGGAPPPORQ 1491
 DB 1457 NSTPMKTPSPVSNITQLOKQKMTQOVNPSPCRCSTREKLTMLPECEBGAGAPPPORQ 1516
 1492 NNADITODLTGNISDYLVTYVQIILAKSLNKNKIVNEFFRGSFSLGVSNTQALPPOEV 1551
 DB 1517 RSTELLQDLTKRINIDFLVKTYPALIRSSLSKPEVNVQRYGGISIG---GKLPVVPIT 1572
 1552 NDAIKQMKRHLKLANDSSADRLNSLGR-----FMIGLDRNNVVKV 1592

Db 1573 GEAIV-----GFLSLGRIMNVSQSPITREASKETLPDLKLEETEDNIKV 1617
 QY 1593 WFNKNGMHAISFLNINNAIRALQGENDSHGITAENHPINLTKOOLSEVALMTTS 1652
 Db 1618 WFNKNGMHAISFLNINNAIRALQGENDSHGITAENHPINLTKOOLSEVALMTTS 1677
 QY 1653 VDLVLSICVIFMSFVSPASFFVFLQERVSRAKHLQFISGKPKVITVLSNVMCMKVYV 1712
 Db 1678 VDAVVAICVIFMSFVSPASFFVFLQERVSRAKHLQFISGKPKVITVLSNVMCMKVYV 1737
 QY 1713 PATVITITFCQOKSYYSINLPVLLALLLYGWSITPLMYPASFEVKIPSTAYVLTTS 1772
 Db 1738 SGLVAVGIFGOKKAYSPENLPALVALLLLYGMAVLPMMYPASFLPDVSTAYVALSC 1797
 QY 1773 VNLFIGINSVATFVLETFDNLNINNDILKSVPLIFPEFCGRLLDMVKQOMADA 1831
 Db 1798 ANLFIGINSVATFVLETFDNLNINNDILKSVPLIFPEFCGRLLDMVKQOMADA 1857
 QY 1832 LEREEENFVSPASFFVFLQERVSRAKHLQFISGKPKVITVLSNVMCMKVYV 1891
 Db 1858 YARFEEHSHANPFWDLIGKMLFAMAVEGVYVFLTLVGRHFFLSQWIAEPKEPIVDE 1917
 QY 1892 DEDVREHORITDGGGNDILEIKELIKYRRKRKPAVDITCIGIPGECFGLGVNGAG 1951
 Db 1918 DDDVAEEFORITDGGGNDILEIKELIKYRRKRKPAVDITCIGIPGECFGLGVNGAG 1977
 QY 1952 KSTFFKMLTGDTVTRGDAPLANKSILSNIEHVQNGYCPQDALTTELIGREHEFFA 2011
 Db 1978 KTTFFKMLTGDTVTRGDAPLANKSILSNIEHVQNGYCPQDALTTELIGREHEFFA 2037
 QY 2012 LLRGVPEKEVGKGMALIKRLGLVYKEKAYSGNKKRISTAMALIGPPVFLDEP 2071
 Db 2038 RLRGVPEKEVGKGMALIKRLGLVYKEKAYSGNKKRISTAMALIGPPVFLDEP 2097
 QY 2072 TTGMDFARRFLMNCALSVYKEGSRVLTSHSMECEALCTRAIVNNGRFGICSYOHL 2131
 Db 2098 TTGMDFARRFLMNCALSVYKEGSRVLTSHSMECEALCTRAIVNNGRFGICSYOHL 2157
 QY 2132 KNRFGDCTIIVLRAGN---PDLKPVQDFGLAPGSGVLEKGRHMLQVLOPSSLSL 2187
 Db 2158 KSKGDDCTIIVLRAGN---PDLKPVQDFGLAPGSGVLEKGRHMLQVLOPSSLSL 2215
 QY 2188 ARFESILSOSKRRHIEDSVSQTLDQYFVNPAKQSDHDLKLSLH 2236
 Db 2216 ARFESILSOSKRRHIEDSVSQTLDQYFVNPAKQSDHDLKLSLH 2261
 RESULT 4
 ABC2_HUMAN
 ID ABC2_HUMAN STANDARD: PRT: 2436 AA.
 AC Q9BZC7;
 DT 16-OCT-2001 (rel. 40, Created)
 DT 16-OCT-2001 (rel. 40, Last sequence update)
 DT 15-JUN-2002 (rel. 41, Last annotation update)
 DE ATP-binding cassette, sub-family A, member 2 (ATP-binding cassette transporter 2) (ATP-binding cassette 2).
 GN ABCA2 OR ABC2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX PubMed=11178988;
 RA Kaminski W.E., Plehler A., Pullmann K., Porsch-Oczurne M., Duong C., Bared G.M., Buchler C., Schmitz G.;
 RT "Complete coding sequence, promoter region, and genomic structure of the human ABCA2 gene and evidence for sterol-dependent regulation in macrophages";
 RL Biochem. Biophys. Res. Commun. 281:249-258(2001).
 CC -1- FUNCTION: Probable TRANSPORTER, ITS NATURAL SUBSTRATE HAS NOT BEEN FOUND YET. MAY HAVE A ROLE IN MACROPHAGE LIPID METABOLISM AND NEURAL DEVELOPMENT.

CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. ABCA SUBFAMILY.
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC
 DR EMBL, AF327705; AAK14335.1; JOINED.
 DR EMBL, AF327658; AAK14335.1; JOINED.
 DR EMBL, AF327659; AAK14335.1; JOINED.
 DR EMBL, AF327660; AAK14335.1; JOINED.
 DR EMBL, AF327661; AAK14335.1; JOINED.
 DR EMBL, AF327662; AAK14335.1; JOINED.
 DR EMBL, AF327663; AAK14335.1; JOINED.
 DR EMBL, AF327664; AAK14335.1; JOINED.
 DR EMBL, AF327665; AAK14335.1; JOINED.
 DR EMBL, AF327666; AAK14335.1; JOINED.
 DR EMBL, AF327667; AAK14335.1; JOINED.
 DR EMBL, AF327668; AAK14335.1; JOINED.
 DR EMBL, AF327669; AAK14335.1; JOINED.
 DR EMBL, AF327670; AAK14335.1; JOINED.
 DR EMBL, AF327671; AAK14335.1; JOINED.
 DR EMBL, AF327672; AAK14335.1; JOINED.
 DR EMBL, AF327673; AAK14335.1; JOINED.
 DR EMBL, AF327674; AAK14335.1; JOINED.
 DR EMBL, AF327675; AAK14335.1; JOINED.
 DR EMBL, AF327676; AAK14335.1; JOINED.
 DR EMBL, AF327677; AAK14335.1; JOINED.
 DR EMBL, AF327678; AAK14335.1; JOINED.
 DR EMBL, AF327679; AAK14335.1; JOINED.
 DR EMBL, AF327680; AAK14335.1; JOINED.
 DR EMBL, AF327681; AAK14335.1; JOINED.
 DR EMBL, AF327682; AAK14335.1; JOINED.
 DR EMBL, AF327683; AAK14335.1; JOINED.
 DR EMBL, AF327684; AAK14335.1; JOINED.
 DR EMBL, AF327685; AAK14335.1; JOINED.
 DR EMBL, AF327686; AAK14335.1; JOINED.
 DR EMBL, AF327687; AAK14335.1; JOINED.
 DR EMBL, AF327688; AAK14335.1; JOINED.
 DR EMBL, AF327689; AAK14335.1; JOINED.
 DR EMBL, AF327690; AAK14335.1; JOINED.
 DR EMBL, AF327691; AAK14335.1; JOINED.
 DR EMBL, AF327692; AAK14335.1; JOINED.
 DR EMBL, AF327693; AAK14335.1; JOINED.
 DR EMBL, AF327694; AAK14335.1; JOINED.
 DR EMBL, AF327695; AAK14335.1; JOINED.
 DR EMBL, AF327696; AAK14335.1; JOINED.
 DR EMBL, AF327697; AAK14335.1; JOINED.
 DR EMBL, AF327698; AAK14335.1; JOINED.
 DR EMBL, AF327699; AAK14335.1; JOINED.
 DR EMBL, AF327700; AAK14335.1; JOINED.
 DR EMBL, AF327701; AAK14335.1; JOINED.
 DR EMBL, AF327702; AAK14335.1; JOINED.
 DR EMBL, AF327703; AAK14335.1; JOINED.
 DR EMBL, AF327704; AAK14335.1; JOINED.
 DR EMBL, AF327705; AAK14335.1; JOINED.
 DR MIM, 600047; ABCA2.
 DR InterPro: IPR003593; AAA_ATPase.
 DR InterPro: IPR003439; ABC_transporter.
 DR Pfam: PF00005; ABC_tran. 2.
 DR ProDom: PD000006; ABC_transporter. 2.
 DR SMART: SM00382; AAA_2.
 DR PROSITE: PS00211; ABC_TRANSPORTER; 1.
 KW ATP-binding; Transport; Transmembrane; Repeat; Glycoprotein.
 FT TRANSMEM 21 40 POTENTIAL.
 FT TRANSMEM 706 728 POTENTIAL.
 FT TRANSMEM 749 771 POTENTIAL.
 FT TRANSMEM 786 808 POTENTIAL.
 FT TRANSMEM 813 835 POTENTIAL.

FT TRANSMEM 850 872 POTENTIAL.
 FT TRANSMEM 892 914 POTENTIAL.
 FT TRANSMEM 1793 1815 POTENTIAL.
 FT TRANSMEM 1846 1865 POTENTIAL.
 FT TRANSMEM 1875 1897 POTENTIAL.
 FT TRANSMEM 1904 1926 POTENTIAL.
 FT TRANSMEM 1988 2010 POTENTIAL.
 FT NP_BIND 1025 1032 ATP (POTENTIAL).
 FT NP_BIND 2088 2095 ATP (POTENTIAL).
 FT CARBOHYD 14 14 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 90 90 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 169 169 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 174 174 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 306 306 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 369 369 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 380 380 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 421 421 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 433 433 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 477 477 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 485 485 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 495 495 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 531 531 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 545 545 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 591 591 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 601 601 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 629 629 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1409 1409 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1497 1497 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1550 1550 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1558 1558 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1613 1613 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1678 1678 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1776 1776 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2055 2055 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SO SEQUENCE 2436 AA: 269971 MW: 966888615DE0ED CRC64;

Query Match 35.9%; Score 4230.5; DB 1: Length 2436;
 Best Local Similarity 39.8%; Pred. No. 4.5e-257;
 Matches 998; Conservative 347; Mismatches 730; Indels 435; Gaps 61;

QY 6 QRLRLMKNLFRRRQTCOLLLEVAMPFLIFLLISVRLSPYEOHECHFPKA-MPSA 64
 DB 6 QLOLLMKNVTLKRRSPVLAFLFIFLVLFFILLGLRQKPTLSVEKVPFYTAAPLTS 65
 QY 65 GTLPVQGITCANNPCRRYPTPEBAPGVVGNENKSIVARLFDARLL-lysokdt 120
 DB 66 GILPVMOSLCPDGGORDEFGFL-----QIANSTVQLLEKIDRVVEEGLNLPDAP 115
 QY 121 SMKDKRVLTLOQIKKSSNMLKLODFLVNDFSG-----FLYHNLSLP 165
 DB 116 SLGSELEALR-QHLEALSAGPTSGSHLDRSTVSFLDSVARNPQELWFLTONLSLP 173
 QY 166 KSTVDMKLRADY---ILAKYFLQGYQHLLTS-LCNGS-----KSEEMI- 204
 DB 174 NSTAQAALLAARDPEVYHLLFEPSSALDSQSLGKGEPMSRLGNGLPMFELLAPA 233
 QY 205 ---QL-----GDOEVSFLGLPKEXLAAB-----RVLRSNMDI 235
 DB 234 LLEQLTCTGSGSELGRLITLVPESSQKALQGYDAVCSQAARARFSGLSAELRNOLDV 293
 QY 236 LKPIILTLNSTSPFSPKELAEA--TKTLHSIGTLAQLFMSKRSMDKROQVMTLVNNS 293
 DB 294 AK-VSQQLGLDAPNGSDSSPOAPPPRRLOALLGLLD-----AQVLOLDVVLIS 341
 QY 294 SSSSTQIYQAVSRIVCGHEGGGLKIKSLNMYEDNNYKALFNGSTEDAEFTYDNTTP 353
 DB 342 ALALLLPQHCIGRTGPPAPASGAGGAN-----GTGAGAVMPNMTADGAPSAALATP 396
 QY 354 YCNDLKNLSSPLSRILIMKALKPLLVG----- 381
 DB 397 ---DLQGGCSAFVQ--LWAGLOPILCGNNRTIEPEALRGNMSSIGFTSKDORNLGLLV 451
 QY 382 -----KILYEDTPATIRQVMAEVNKTFOELAVFHDLEGMWEELSPK IWTMENSQEMD 434

DB 452 HLMTSNPKIILYAPAGSEVDRIILKANETFAFVGWVTHYAQVWLNINISAEIRSELEGRLOO 511
 QY 435 LVRLMLDSRDNDHFEQOOLGDLMTAODIYAFIAFLAKHEDQSSNGSYWTRAEFNTN-- 492
 DB 512 HLRLV-----OOYAEVLRPE---ALNLSIDELPRALRDNFS 547
 QY 493 -----QALRTIS-----RMECVNLNKLPIETAEVWLINKSME--LDERKFWAG 535
 DB 548 LPSGALLQOOLDTIDNAACGMIOKMSKVSVDIFKGFPEDESIYVTLNQAQONVTVFAS 607
 QY 536 IVFTGITPQSTLELRPHVKKIIRMDIDNVERTNKTIDGYWDPGRPADFEEDRYWVGGEAV 595
 DB 608 VIFQPTKQDS--LRPHVYKIRONSFEKTEKTEIRRAWRGPPWG---GRFYLGFTW 662
 QY 596 LODVQOAIIRVLGTE-KKGVYQMOQMPYCYDDIFLRYMSKSMPLMTLMAIYSVAV 654
 DB 663 IDMMERAIIDFVGDVDPQSVQMPRYPCYTRDDPLFVLEHMMPLQWISVSVAM 722
 QY 655 IIKGIYEREARLKETMRIMGDINSILMFSEFISLPLVLSAGILVYLKGLNLPYSD 714
 DB 723 TIQHIYAEKHEKRLKEVMKTMGLNNVHVAWMTGTFVOLDSIVTALTKYGOVLMHSH 782
 QY 715 PSVVEVPLSYFVNTTLOCFILSTLESRAVLAACGGIIFLYLPLVLC-----VAMOD 769
 DB 783 VVIIMFLAYAVATIMFCFLSVLSKAKLASACGGIIFLYLPLVLC-----VAMOD 842
 QY 770 YVGFTLKFIFALSIPAFAFGCEYFALFEEOGIGVQMDNLFESPEVEDGPNLTISML 829
 DB 843 ITAFK-KCIASLMSTAFAGISGYFALYEVAGVGIOMHTFSQSPVEGDFNLLAVTLM 901
 QY 830 FDFPLGYMTWYIEAVFPGQYGIIPRWYFPCIKSYNFG---ESDEKSHPSGNOKRMS-- 884
 DB 902 VDAVVYGIILWYIEAVFPGQYGIIPRWYFPCIKSYNFG---ESDEKSHPSGNOKRMS-- 961
 QY 885 ---ETC-----MEEPETHLKIGVSIQNLKVVYRDGKVAVDGLALNFYEGOTT 929
 DB 962 EEDACAMESRREFETGMEEPETHLPLVVCVDPKLLKVKYKDKOKLALNKLSINLYENOV 1021
 QY 930 SFLGNAGKTTMSILTLGPPTSGTAYILGKODISEKSTTQNLGCVQPHNVLFMDLT 989
 DB 1022 SFLGNAGKTTMSILTLGPPTSGTAYILGKODISEKSTTQNLGCVQPHNVLFMDLT 1081
 QY 990 VEEHIMFARLKGISEKHVAKEMOMALDVLPSKLSKTSQLSQSGMOKRKLVALAFVG 1049
 DB 1082 VEEHIMFARLKGISEKHVAKEMOMALDVLPSKLSKTSQLSQSGMOKRKLVALAFVG 1140
 QY 1050 GSKVYILDEPTAGVDPYSSRGIMWELLKYRGRTIILSTHMDADVLDRALAIISHGL 1109
 DB 1141 GSRALIDEPAGVDPYARRAIMDLIKYKPGRTILSTHMDADVLDRALAIISHGL 1200
 QY 1110 CCVGSFLKNOGLGYTLTVKKDYESLSSCNSSSTVYLKEDSVSSSDAGLGS 1169
 DB 1201 KCGSPFLKGYGGVGYTLTKRPAEPG-----GPOFGLAS 1238
 QY 1170 DHESDPTLDVS--AISNLIRKHVSEARLEDIGHETLVLPYPAKEGAFVLEPHED 1227
 DB 1239 SPFGARPLSSGSELQVQFIRKHVASCLVSDTSTELSYILPSEAAKGAFFERFOHLER 1298
 QY 1228 RUSDLGISTSYGISETTLEIFLKAVE-----SGVNA-ETSDQTL 1267
 DB 1299 SLDALHLSSFGIMDTTLERVLKVSIEDOSLENSADVAKESRKVDLFGACBPASGEBHAG 1358
 QY 1268 -----ARRNRA--FGDKQSCILAPFTEDDAADPNDSDIDPESR 1303
 DB 1359 NLRGSELVQSOASLOSASVSARGDEAGCTDYGYRPLF--DNQDPD--NVSLOEV 1415
 QY 1304 ETDLLSGMDGKGSYQYKGLLQOOFVALLKRLIARRSKRGPTAOLVPLAVFVCLALV 1363
 DB 1416 EBAVALSRV--GQGSRKLDGWLKVRQFHGLLVKRFICARRNSALPESOLTLPAFFVCVAMT 1474
 QY 1364 FSLIVPFGKYPSTLELQPMWYNQYV-----FVSNDAPE-----DTGTELELNAL 1408


```

OY 1731 SSNTLPVALALLLLXGNSITPLMPASVFEKIPSTAYVILTSVNLFTGINSVATVELEL 1790
DB 1869 SPTNPAPVASTLFLIGMSITPLMPASVFEVPSAVVFLVILNLFIGITATVATFLLQ 1928
OY 1791 FV-DNKLNNDIKLSVLPFPHFCIGRLIDYKNOAMDALERGE-NREVSPLSMWL 1848
DB 1929 FHDHDLKVVNSYLSKCLFEPVNLGIGLMEAMAYNEYINETYAKIGQFPMKMSPEEMDI 1988
OY 1849 VGRNFMAAVGVFEFLITVLIQVREFTRPPVNAKISPLNDEDEDVRRERQRIIDGCGQ 1908
DB 1989 VTRGLVANTVAGVGFELTLMQVNFLOPQRLPVSTKPEVD-DVDAVSHRQVNLGDDAD 2047
OY 1909 NDLEIKELTKTYRRK---RKPADVRLICVIGIPCECGGLGVAGKASSFFKMLTGDTTV 1965
DB 2048 NDMVKIEMITKVYKSRKIGRLAVDRCLCVRGECGGLGVAGKASRTSFFKMLTGDEST 2107
OY 1966 TRGDAFLKNSIISNIEHVQNMKYQCFATITELLTGREHVEFALLRGVPEKEVKG 2025
DB 2108 TGGAFVNGHSHVLDLQVQSLGCPQFALDELTAHRLQLYRLKGIIPMKDEAGV 2167
OY 2026 EMATRKGLIVKYGKRYAGNSGKRRKLTAMALIGPPVFLDEPTTGMCKARFLMN 2085
DB 2168 KMALEKELTKYADKRPAGTYSGKRRKLTALITPAPFPLDEPTTGMCKARFLMN 2227
OY 2086 CALSVVEGSHVLTSHSMECALCTRMALVNGRFLGSHVQHLKNRFGDGYTVIRI 2145
DB 2228 LILDLITGSHVLTSHSMECALCTRLAIVNGRFLGSHVQHLKNRFGDGYTVIRI 2286
OY 2146 AGSNPDLKPVDFGLAFGSVLEKERNMLQYOLPSSLSLARFSLISQSKKLHED 2205
DB 2287 TKSSQVWKDVYRFPFNRPPEAMLEKHHKVVYOLKSEHISLADVFSKMEGVVGLIED 2346
OY 2206 YVSQOTTLOVFNFAKQSDSD 2227
DB 2347 YVSQTTLLDNVFNFAKQSDSD 2368

```

RESULT 6
ABC3_HUMAN
ID ABC3_HUMAN STANDARD: PRT: 1704 AA.
AC 099758: 092473: Rel. 39, Created
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ATP-binding cassette, sub-family A, member 3 (ATP-binding cassette transporter 3) (ATP-binding cassette 3) (ABC-C transporter).
GN ABC3 OR ABC3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Thyroid carcinoma;
RX MEDLINE=9632608; PubMed=8706931;
RA Klugbauer N., Hofmann F.;
RT "Primary structure of a novel ABC transporter with a chromosomal localization on the band encoding the multidrug resistance-associated protein";
RL FEBS Lett. 391:61-65(1996).
[2]
RP SEQUENCE FROM N.A.
RX MEDLINE=9719225; PubMed=9027511;
RA Connors T.D., Van Raay T.J., Petry L.R., Klinger K.W., Landes G.M., Burn T.C.;
RT "The cloning of a human ABC gene (ABC3) mapping to chromosome 16p13.3";
RL Genomics 39:231-234(1997).
CC - FUNCTION: MAY BE A TRANSPORTER, ITS NATURAL SUBSTRATE HAS NOT BEEN FOUND YET (BY SIMILARITY). MAY ACT AS AN EFFLUX PUMP FOR CHEMOTHERAPEUTIC DRUGS.
CC - TISSUE SPECIFICITY: HIGHLY EXPRESSED IN LUNG, FOLLOWED BY BRAIN,

```

CC PANGREAS, SKELETAL MUSCLE AND HEART. WEAKLY EXPRESSED IN PLACENTA, KIDNEY AND LIVER. ALSO EXPRESSED IN MEDULLARY THYROID CARCINOMA, CELLS (MTC) AND IN C-CELL CARCINOMA.
CC - DOMAIN: MULTIFUNCTIONAL POLYPEPTIDE WITH TWO HOMOLOGOUS HALVES, EACH CONTAINING AN HYDROPHOBIC MEMBRANE-ANCHORING DOMAIN AND AN -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. ABCA SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U78735; AAC0967.1; -
DR EMBL: X97187; CAA65825.1; -
DR Genew: HGNC:33; ABCA3.
DR MIM: 601615; -
DR InterPro: IPR003593; AAA_Atpase.
DR InterPro: IPR003439; ABC_transportr.
DR Pfam: PF00005; ABC_tran.2.
DR ProDom: PD000006; ABC_transportr; 2.
DR SMART: SM00382; AAA; 1.
DR PROSITE: PS00211; ABC_TRANSPORTER; 1.
DR KATP-binding; Transport; Transmembrane.
FT TRANSMEM 22 42 POTENTIAL.
FT TRANSMEM 249 269 POTENTIAL.
FT TRANSMEM 307 327 POTENTIAL.
FT TRANSMEM 344 364 POTENTIAL.
FT TRANSMEM 373 393 POTENTIAL.
FT TRANSMEM 405 425 POTENTIAL.
FT TRANSMEM 447 467 POTENTIAL.
FT TRANSMEM 925 945 POTENTIAL.
FT TRANSMEM 1100 1120 POTENTIAL.
FT TRANSMEM 1144 1164 POTENTIAL.
FT TRANSMEM 1183 1203 POTENTIAL.
FT TRANSMEM 1213 1233 POTENTIAL.
FT TRANSMEM 1245 1265 POTENTIAL.
FT TRANSMEM 1306 1326 POTENTIAL.
FT NP_BIND 566 573 ATP (POTENTIAL).
FT NP_BIND 1416 1423 ATP (POTENTIAL).
FT CONFLICT 36 36 P -> S (IN REF. 2).
FT CONFLICT 196 196 L -> P (IN REF. 2).
FT SEQUENCE 1704 AA; 191367 MW; AF0098DAF7A04F5F CRC64;

```

Query Match 22.4%; Score 2642.5; DB 1; Length 1704;
Best Local Similarity 35.7%; Pred. No. 1.7e-157;
Matches 639; Conservative 300; Mismatches 523; Indels 327; Gaps 48;

```

OY 534 AGIYFTGJTPGSI-EIPHNKYYKIRMDINVERTN-----KIKDGYW----- 574
DB 131 AAVFEPHNPMSKEDPLAVAKYLR-----SYFRNNMTOTSGFFLKEGHTTSLRPL 187
OY 575 --DPRADPFED---MRVYMGFAVLDVVEQALIRVLTGE-----KKGIVYQOMRY 624
DB 188 FPNRGPRLTSPDGGEGYVIREGFLAVQHAVDRALMEYADATQOLFQRLTVYTKRPY 247
OY 625 PCYVDITFLRVMSRMPLEMTLAWISVAVIIGIYVEKARLKEKTRIMGLDNLMSLWFS 684
DB 248 PPTLADFLVAYOYQPLRLLLSFYTTALTTLARAVQEKERLKYEMMGSLMHSWA 307
OY 685 WFTSLIPLVSAGLLVILKLG-----NLDPYSDPSVVFPLSFVAVNTLQCFPLSTL 739
DB 308 WFLFLFLILLIASFWTLFLCYKVPNNAVLSRSDPSLVLAFLCFAISTISFSFWSTF 367
OY 740 ESRANIAAACGGITFTTLVPLVLCVAMQDYVGLTKTFASLSPVAFGFCYFALFEE 799
DB 368 FSRANAAAFGGFLVFFTYIIPYFVAPRYNMTLSQKCSLLSNVAMAGAOILGFEA 427
OY 800 OGIGVQMDLFPSSPVE-EDGFNLVTSISMMLFDPLVGVMTWYIEAVPQOYGIPTRWYF 858

```

```

Db 428 KGMGIOMRDL-SPVNVDDPCFGVGLMLLDSVLGLVITWMEAVPQGFVQPMWF 486
OY PCTKSTWGE-----ESDEKSHGSKNRMSELICMEEEPHLTKLGSIONLVKYEDGM 912
Db 487 FIMPSTWCKPRAVAGKEEEDSDP---EKALRNEYFEKPEBDIAGIKIKILSVFRVGN 543
OY 913 K--VAVDGLALFNYEGQITSPFLGNGAGKTTTMSLTGLFPPTSGTAVYILCKDIRSEMT 970
Db 544 KDBAAVRDNLMLYEGQITVLLGNGAGKTTTMSLTGLFPPTSGTAVYILCKDIRSEMT 603
OY 971 IRONLCVCPQHNVLPDMLTVEEHIWFTARLKGSLSEKHYKAEQMDALDVGLPSSKLSKT 1030
Db 604 IRKSLGLCPQHNLFDNLVLAHELYFAQLKGLSRQCKPEEVKQMLHTIGL-EDKWNRS 662
OY 1031 SOLSGMOKRSLVALAFVGSKVVLDEPTAGVDPSRRGIWELLKRYGROGRTIILSTHN 1090
Db 663 RFLSGMRRLKSLGLALNGSKVLLDEPTSGMDAISRAITMDLQKRSRTIYLTTHF 722
OY 1091 MDEADVLGDRIALIISHGLCCVSSSLFLKNQIGTYVYLLVKKDYESSISGRSSSTVS 1150
Db 723 MDEADLLGDRIALIMAKGELQCCSSSLFLKQYAGYHMTLVKE-----P 766
OY 1151 YLKEDSVSSSSDAGLSGHESDITLTIDVSAISNLIRKHVSEARLVEDIGHETVLPY 1210
Db 767 HCNED-----ISQLVHHHVNATLESAGAEISFILPR 800
OY 1211 EAKKEGAFELFHEIDRLSDLGISYGISETTLEIEPLKVAE-ESGVDAETSDGTLPA 1268
Db 801 ESTHR--FEGFLFKLEKKEKELGASFGASITMEVEVPLRVGLVDSMDIOAIO--LPA 856
OY 1269 ---RRNRAGF--DKOSCLRPTEDDADPND---SDIDPERETDLSGMDKGSYOV 1319
Db 857 LOYOHERRASDAVDSMLC-----GAMPDSGICALIEEETAVAKLWTLG----- 901
OY 1320 KGMKLTQOQFALIMKRLLIARSRKGFPAOIVLPAVFCIALVFSLIAPPGKYPSEIL 1379
Db 902 ---ALHCOQEFAMFLKKAANSWREKMWAAOVVPLTCVTLALL----- 942
OY 1380 QPMWYNEQYFVNSDABEDTGTLELLNALTFRPGFTGRCMEGNPFDPTQOAGEEWTTA 1439
Db 943 -----AINYSELFDPPML--RLTLG-----EYGR 966
OY 1440 PVPOTIMLPONGWMTQONSPACQSSDKIKKMLPVCPPGAGGLPPQKONTADILAD 1499
Db 967 VPFVSPGTSGOGLSHLKDALQAGS-----QPREVLAD 1003
OY 1500 LTRNISDYLVKTYVQIILAKSLKNKIWNEFRY---GSEGLSVNTQALLPSQEVNDAL 1555
Db 1004 L-----EEFLI-----FRASVEGGGFN----- 1020
OY 1556 KOMKHLKLAKSSADRLNLSLGRPMGTGTRNNVYKWFNNKGMHAISFLVYNNALIR 1615
Db 1021 -----ERCL--VAASFBDVGERTYVNALENNQAHSPATLAAVYDNLLEK 1063
OY 1616 ANLQGENPNSHGITAFNHP-----LNTLQOLSEVALMTSTVDVLSVIVFAMSEVPA 1670
Db 1064 ---LLCG---PHASIVYNSFPQPRSLAQAKQFNE---GRGCPDIALNL-LFMARLAS 1113
OY 1671 SEVVELLQERVSRAKHLQFISGVKPVYIWSNFWDMCNVYVATVYIILFCIQOKSYV 1730
Db 1114 TFSILAVSRRAVQAKHVGQVSGVHASFWSLMLMDLISFILPSILLVYKADVRAFT 1173
OY 1731 SSTNLPVALLLLLYGMSTTLPMTYSPASVFEKIPSTAVYVLLSVNLFJIGINSVATFVEL 1790
Db 1174 RQGHMADTLLLLYGMALITPLMTLMNEFFLGAATATYRLLITFNLSI-----ATFLMVT 1229
OY 1791 ---FTDNKLNINDILKSVLFLPHFCGLGLIDVKN-----QAMADALERRG- 1836
Db 1230 IMIIPAVKLEELSKTLDHVFVLPHNCLGMAYSEYENWETFRYCTSESEVAHCKKXNI 1289
OY 1837 ---ENRNPVPLSMDL--VGRNLFMAAVEVFEITVLIOYRFFIRPPVANKL----- 1885
Db 1290 QYOENFY---AMSAPGVGRFVASMASGCAVLLLFLETINLQRLKGLICALRRKRL 1345

```

```

OY 1886 -----SPUNDEDEDYRRERORIIDGGGONDI---LEIKELTKYRRKKRP--AVDRICV 1934
Db 1346 TELYTRMPLPEDDVADERTRIILAPSPDSLHTPLIKELSKY-ROKRPPLAVRLSL 1404
OY 1935 GTPCECFGLGVNGAGKSTFEKMLTGDTYTRGDAFLNKNLSINIEHYQNGYCPQF 1994
Db 1405 AVQKECEGLLFQNGAGTTFTKMLTGEEESLTGSDAFVGHRRISDVGKVRORIGYCPQF 1464
OY 1995 DATTELLTGREHVEFPALLRGVPEKEVKVGBMAIRKIGLVKYEKXAGANSQGNKRLS 2054
Db 1465 DALDHMTGREMIVYARLGRIPERHIGACVENTLGLLEPHANKLVRYTSGGNKRLS 1524
OY 2055 TAMALIGPPPVVFLDEPTGMDPKARREFLMNCALSVYEGRSVLTSHSMECEALCTRM 2114
Db 1525 TGIALIGEPVAVFLDEPTGMDPKARREFLMVLAARREGKAILITSHSMECEALCTRL 1584
OY 2115 AIMVNGRFRCLGSVQHLKKNRFGDGYTYIVRI--AGSNPLKPVODFGLAFPGSVLKEKH 2172
Db 1585 AIMVQGFCLGSPQHLKSKFGSGYSLRAKVGSEGOBALLEEFKAFVDLTTPGSLVEDH 1644
OY 2173 RNMLQYQLPSSLSLARIFSILOSCKRLHIEDYSVSTOTPDVFNPA 2221
Db 1645 QGMVHYHLPGRLDLSMAKVPGLLEKAKETGVDDYSQISLEQVFLSFA 1693

```

```

RESULT 7
CED7_CAEEL STANDARD: PRT: 1704 AA.
ID CED7_CAEEL STANDARD: PRT: 1704 AA.
AC P34358; P34359; 076287;
DT 01-FEB-1994 (Rel. 28, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ABC transporter ced-7 (cell death protein 7).
GN CED-7 OR C48B4.4.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_Taxid=6239;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM C), FUNCTION, AND MUTAGENESIS OF LYS-586;
RP GLU-639 AND LYS-1417.
RC STRAIN-Bristol N2;
RC MEDLINE=98297348; PubMed=9635425;
RA Wu Y.-C., Horvitz H.R.;
RA "The C. elegans cell corpse engulfment gene ced-7 encodes a protein
RT similar to ABC transporters.";
RL Cell 93:951-960(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-Bristol N2;
RC MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Cratton M., Dear S., Du Z., Durbin R., Hawkins T., Hillier L., Jier M.,
RA Fulton L., Gardner A., Green P., Kershaw J., Kistner J., Laister N.,
RA Johnston L., Jones M., Kershaw J., Lloyd C., Mortimore B., O'Callaghan M.,
RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Snowkneen R.,
RA Sims M., Smaildon N., Smith A., Smith M., Sonhammer E., Staden R.,
RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
RA Watson R., Watson A., Weinstock L., Wilkinson-Sproat J.,
RA Wohldman P.;
RA "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
RN [3]
RP REVISIONS, AND ALTERNATIVE SPLICING.
RA Durbin R.;
RA Submitted (OCT-2000) to the EMBL/Genbank/DBJ databases.
CC -i- FUNCTION: Functions in the engulfment of cell corpses during
CC embryonic programmed cell death to translocate molecules that
CC mediate homotypic adhesion between cell surfaces of the dying and
CC

```

CC	engulfing cells.
CC	-i- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC	-i- ALTERNATIVE PRODUCTS: 3 isoforms; a, b and c (shown here); may be produced by alternative splicing.
CC	-i- TISSUE SPECIFICITY: Ubiquitous in embryos. Expressed in larval gemmule precursors. Expression in larvae and adults is seen in amphid sheath cells, pharyngeal-intestinal valve and phasmid sheath cells. Low levels of expression are also seen in gonadal cells.
CC	-i- DOMAIN: Multifunctional polypeptide with two homologous halves, each containing a hydrophobic membrane-anchoring domain and an ATP binding cassette (ABC) domain.
CC	-i- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. ABCA SUBFAMILY.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - CC the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial CC entities requires a license agreement (See http://www.isb-slb.ch/announce/ or send an email to license@isb-slb.ch).
CC	-----
DR	EMBL; AF049142; AAC24116.1; -
DR	EMBL; Z29117; CA82384.2; -
DR	EMBL; Z29117; CA82383.2; -
DR	EMBL; Z29117; CAC42271.1; -
DR	PIR; S40724; S40724.
DR	PIR; S40725; S40725.
DR	WormPep; C48B4.4a; CE24856.
DR	WormPep; C48B4.4b; CE24857.
DR	WormPep; C48B4.4c; CE27867.
DR	InterPro; IPR003593; AAA_Arpsase.
DR	InterPro; IPR003439; ABC_transporter.
DR	Pfam; PF00005; ABC_tran.2.
DR	ProDom; PD000006; ABC_transportr; 2.
DR	SMART; SM00382; AAA; 2.
DR	PROSITE; PS00211; ABC_TRANSPORTER; 2.
KW	ATP-binding; Transport; Transmembrane; Repeat; Glycoprotein;
KW	Alternative splicing.
FT	TRANSMEM 24 44 POTENTIAL.
FT	TRANSMEM 256 276 POTENTIAL.
FT	TRANSMEM 306 326 POTENTIAL.
FT	TRANSMEM 335 355 POTENTIAL.
FT	TRANSMEM 436 456 POTENTIAL.
FT	TRANSMEM 963 983 POTENTIAL.
FT	TRANSMEM 1126 1146 POTENTIAL.
FT	TRANSMEM 1176 1196 POTENTIAL.
FT	TRANSMEM 1201 1221 POTENTIAL.
FT	TRANSMEM 1234 1254 POTENTIAL.
FT	TRANSMEM 1311 1331 POTENTIAL.
FT	NP_BIND 580 587 ATP (POTENTIAL).
FT	NP_BIND 1411 1418 ATP (POTENTIAL).
FT	CARBOHYD 126 126 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD 145 145 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD 359 359 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD 421 421 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD 427 427 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD 481 481 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD 678 678 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD 727 727 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD 899 899 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD 986 986 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD 1012 1012 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD 1045 1045 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD 1597 1597 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD 1632 1632 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	VASPLIC 992 993 MISSING (IN ISOFORM A).
FT	VASPLIC 496 508 MISSING (IN ISOFORM B).
FT	MOTAGN 486 586 K->R: CELL CORPSES NOT ENGULFED.
FT	MOTAGN 639 639 E->G: CELL CORPSES NOT ENGULFED.
FT	MOTAGN 1417 1417 K->R: SOME CELL CORPSES NOT ENGULFED.
SO	SEQUENCE 1704 AA: 191411 MW: 1750240B24507CFC CRC64;

Query Match 13.0%; Score 1538.5; DB 1; Length 1704;
Best Local Similarity 24.9%; Pred. No. 4.3e-88;
Matches 484; Conservative 347; Mismatches 675; Indels 441; Gaps 59

QY	458	WTAODIV-----	-AFLAKHPEDVOSNSGVY-----	TYREAFNFEN-----	492
		: : :	: : :	: : :	
Db	24	WTIFELLIPCLLOPIVLYVKNADHHS	SPRENTIDNQVAGTVEVDFLESNFTKPYKRW	83	
QY	493	-----QAIRTI-----	SFPMECVNLNKI-----	EPJAEVWLINKS	523
		: : :	: : :	: : :	
Db	84	CLRSDDVVGYSKDAARKATVYDLMKKEFAERFOSAKLIKLSVKNESSEODLT	---VLRND	140	
QY	524	MELIDE-----	RKFMAGIVFTGLIPGSELPHVHKYKIRBDIONERTKIDQGYDDP	578	
		: : :	: : :	: : :	
Db	141	LPMNEFFCALINSTAGVAVDEVDVYTNKL	-----NTRILGKTPEETMNLTFSTYNPY	196	
QY	579	RADPFEDM-----	KYWGGAFLQDDVVEQALIRVLGTERTKTYGVMQMPYPCVD	629	
		: : :	: : :	: : :	
Db	197	SSGRYSRI PSSPPYMTSAFLTFQHAIESFSLSVSGARDPLITTLGDEPREPKYSSVA	256		
QY	630	--DIFELWMSRMPLEMTLAMIYSVAVII	KGYVEKAEALKEIMRMIGLONSTLFWMT	687	
		: : :	: : :	: : :	
Db	257	FIDF-----	PEIMAFYTFINVHITREIAENHA-VKPYLWAGMISTEMFAAHV	307	
QY	688	SSLIPLYVSGLLVILIKGNLPLYS	DPSPVVEVFLSVFAVNYIILOCFILSTFSRANLAA	747	
		: : :	: : :	: : :	
Db	308	MAELKFEVI--FLCSIIPIPTFVEMFSPALIVYLMVGLAVIGAVAFSPFNNTSAI	365		
QY	748	ACGGIIYFTLYLPYCLVCMADQVGTLLKIFASL	--LSV-----	ANGCEYFAL	796
		: : :	: : :	: : :	
Db	366	K-----	ALVAMGAMIGISTYKLRPELLOJISCFLYLNINGAFALVAEASD	412	
QY	797	FEEQIGVQVOMDLF--ESPVEEDGFNLTTISIML	FDTEFLGYMTWYIEAV-FPGQYIPR	854	
		: : :	: : :	: : :	
Db	413	YMRERELNLNFMENSSLSH--	FSIGMALVMYVIDLIMSIGALVYDHITSADFSLRT	469	
QY	855	PWFEPCTKSYVNGEESDESHPG	-----SNOKRMSIEMEEPR--	THL---	895
		: : :	: : :	: : :	
Db	470	LEPFE-----	APEDDEMOTDQVTAQNTREINQVNRNVRSRSDMEIOMNMASTLMPN	522	
QY	896	-----	KLGVSIOMLVKYRGMGVAAVADGALINYEBOISFLIGN	935	
		: : :	: : :	: : :	
Db	523	ADSDSLLEGSTPADGADPTARADII	IVRLNVKIMWTTGRAYADGISLRAVQCSILLIGN	582	
QY	936	GAGKTTMILGLFPPPSGTAYILGKQIRSEDMSTIRONLCVCPDHNVLFDMLTVEEHIW	995		
		: : :	: : :	: : :	
Db	583	GAGKSTTSSITAGIIRPTNGRIITCGYVNGEPEBETRRHIMCPOYNPDIQDLTVSEHLK	642		
QY	996	EYARLKLSEKHVAKEMQMALDVGLPSSKSKTSIOLSGGMOKLSVALAFVGSQVVI	1055		
		: : :	: : :	: : :	
Db	643	LYVGLKGAERKDFKQDMKRLLSDVKL	-DFKNEKAVNLSGGMKKKLCVCMALLIGDSFVYL	701	
QY	1056	IDEPFAGVDPYSRKRIMELLKRYROGRTI	IISTHNMADVLDGRIAIISHGLKCCVGS	1115	
		: : :	: : :	: : :	
Db	702	IDEPFAGMDPPAKDOVOVLVERKANRT	ILITTHYHDEARLISGWFIMSHGLVASGTN	761	
QY	1116	LELKLQGTGYLLTKVDVDESSLS	CRNSSSTVSYLKKEDSVQSSDMLGSDHESPT	1175	
		: : :	: : :	: : :	
Db	762	QYLKQKFGTGYLLTV	-----	LHNGDK	784
QY	1176	LTIQVSAISNLIRKHVSEARLVEDIG	HELTLYVLYPEAKEGAEVELFHEID	1226	
		: : :	: : :	: : :	
Db	785	RKMAV-ILTDVCHYVEAEERGEHNGQOILE	ILPEARKKE--FVPLFQALAEIADONKNTS	841	
QY	1227	-----	DRJLSDGJISYSGISETTLEIFILKAAEESGVDAEFTSDGLPARNRRAE	1275	
		: : :	: : :	: : :	
Db	842	NVEDNMENITLKSQATILEMBSFG	SLNTLLEQVFTTIGDK-----	VDAAIASRONSRTS	894
QY	1276	GDKQSCILRPTEDDADPNDDIDPES	RETRDLSGMDGSGSYQVGMKLTQOQFALLMK	1335	
		: : :	: : :	: : :	
Db	895	HNSRNASEPSLKPAGYDQTSSTK	SADSYQK--LMDSQANGP-EKSSVAKMAVQFISIMK	951	
QY	1336	RLTIARSRKGFQAIYILPAVPCIALV	SLVYPPGKATPSLELDQFMATNFOEYTVESDA	1395	

ID NOD1_RHISN STANDARD: PRT: 343 AA.
AC P55476;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Nodulation ATP-binding protein I.
GN NOD1 OR Y4HF.
OS Rhizobium sp. (strain NGR234).
OC Plasmid sym pNGR234a.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OX Rhizobiaceae; Rhizobium.
NCBI_TaxID=394;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97305956; PubMed=9163424;
RA Freiberger C.A., Fellay R., Bairoch A., Broughton W.J., Rosenthal A.,
RA Perret X.;
RT "Molecular basis of symbiosis between Rhizobium and legumes.";
RT Nature 387:394-401(1997)
CC -1- FUNCTION: FORMS, WITH NODJ, A MEMBRANE TRANSPORT COMPLEX INVOLVED
CC IN THE NODULATION PROCESS. IT PROBABLY EXPORT A MODIFIED BETA-1,4-
CC LINKED N-ACETYLGLUCOSAMINE OLIGOSACCHARIDE.
CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
CC -1- SUBCELLULAR LOCATION: Membrane-associated.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: AE000076; AAB91694.1;
DR InterPro: IPR003593; AAA_ATPase.
DR InterPro: IPR003439; ABC_Transport.
DR Pfam: PF00005; ABC_tran; 1.
DR ProDom: PD000006; ABC_transport; 1.
DR SMART: SM00382; AAA; 1.
DR TIGRfams: TIGR01288; nod1; 1.
DR PROSITE: PS00211; ABC_TRANSPORTER; 1.
DR KMW Modulation; Membrane; ATP-binding; Transport; Plasmid.
FT NP_BIND 77 84 ATP (POTENTIAL).
SQ SEQUENCE 343 AA; 37917 MW; F49A7EC56E099A33 CRC64;

Query Match 3.2%; Score 382.5; DB 1; Length 343;
Best Local Similarity 35.1%; Pred. No. 1.3e-16;
Matches 91; Conservative 46; Mismatches 101; Indels 21; Gaps 6;

QY 872 EKSHGSGN-----QKRMSEICMEEPHHLKLGISTQNLVAVYRDGMKAVAVDGLALNFEYEG 926
DB 24 QKCHHSHADNSLSRSKSDVAIE-----LTNVSRSYGD--KVVVDQLSFTITSG 70
QY 927 QITSEFLGNAGKTTTMSILTGLFPPTSGTAVILGDKIREMSTIQNLGVCPOHNVLPD 986
DB 71 ECFGLLPNGAGKSTVSLVGLAPDEGITVVLGEVPAARARLARISRGVVPQPFOTDTR 130
QY 987 MLTVEEHIMFYARLKGISEKHVAKEMQNALVDGLSPSKLSTKSTQSGMOKRLISVALA 1046
DB 131 EFTAREMLVFGKFGHTRLELEAIPPL-LDFARLESKADVPVAVQLSGMGRLTLTAC 189
QY 1047 FVGGSKVVLIDEPAGVDPYSRGIVELLLK-YROGRITILSTHMDADVGLRIATIS 1105
DB 190 LINDPOLLIDEPDTGDPHARHLIWERLRLSLALGKTLTLTHMEERADRLCDRLCYIE 249
QY 1106 HGLTCCVG-SSEFLKNOGL 1123
DB 250 HGKRIYGRPHALIDEOIG 268

RESULT 10
NOD1_RHIS3
ID NOD1_RHIS3 STANDARD: PRT: 304 AA.

AC P72335;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Nodulation ATP-binding protein I.
GN NOD1.
OS Rhizobium sp. (strain N33).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OX Rhizobiaceae; Rhizobium.
NCBI_TaxID=103798;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96303537; PubMed=8755627;
RA Cloutier J., Laberge S., Prevost D., Antoun H.;
RT "Sequence and mutational analysis of the common nodBCJ region of
RT microsymbiont of both arctic and temperate legumes.";
RT Mol. Plant Microbe Interact. 9:523-531(1996).

CC -1- FUNCTION: FORMS, WITH NODJ, A MEMBRANE TRANSPORT COMPLEX INVOLVED
CC IN THE NODULATION PROCESS. IT PROBABLY EXPORT A MODIFIED BETA-1,4-
CC LINKED N-ACETYLGLUCOSAMINE OLIGOSACCHARIDE.
CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
CC -1- SUBCELLULAR LOCATION: Membrane-associated.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: U53327; AAB16898.1;
DR HSSP: O58663; 168H.
DR InterPro: IPR003593; AAA_ATPase.
DR InterPro: IPR003439; ABC_Transport.
DR Pfam: PF00005; ABC_tran; 1.
DR ProDom: PD000006; ABC_transport; 1.
DR SMART: SM00382; AAA; 1.
DR TIGRfams: TIGR01288; nod1; 1.
DR PROSITE: PS00211; ABC_TRANSPORTER; 1.
DR KMW Modulation; Membrane; ATP-binding; Transport.
FT NP_BIND 38 45 ATP (POTENTIAL).
SQ SEQUENCE 304 AA; 33698 MW; 7C6A33B0364CC614 CRC64;

Query Match 3.1%; Score 366; DB 1; Length 304;
Best Local Similarity 35.0%; Pred. No. 1.2e-15;
Matches 77; Conservative 54; Mismatches 85; Indels 4; Gaps 3;

QY 896 KLGVSIGNLVKVVYRDGMKAVAVDGLALNFEYEGQITSEFLGNAGKTTTMSILTGLFPPTSG 955
DB 3 KVAIDLAVGKRSFGD--KLNVNGLSTFVASGEFGTGLGNGAGKSTIARMLGKGVPPAG 60
QY 956 TAYILGDKIREMSTIQNLGVCPOHNVLPDMLTVEEHIMFYARLKGISEKHVAKEMQ 1015
DB 61 KITVYGLPEVARSRLARKSTIGVVPQFNDQEFVRENILVGVFGSTKIK-EVIPS 119
QY 1016 ALDVGLPSSKLSKTSQLSGGMOKRLISVALAVGSKVVLIDEPAGVDPYSRGIVE-L 1074
DB 120 LLEFARLESKADARVGEISGGMKRLITAKALINDPOLLIDEPDTGDPHARHLIWERL 179
QY 1075 LKVRGRTITLSTHMDADVGLRIATIS 1114
DB 180 RFLNARGKTLTLTHMEERADRLCDRLCVLEHGKRLAGS 219

RESULT 11
YADG_ECOLI
ID YADG_ECOLI STANDARD: PRT: 308 AA.
AC P36879;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)

```

DE      Hypothetical ABC transporter ATP-binding protein yadG.
OS      yadG OR B0127.
CC      Escherichia coli.
CC      Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
CC      Escherichia.
XX      NCBI_TaxID=562;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=K12 / W3110;
RX      MEDLINE=94261430; PubMed=8202364;
RA      Fujita N., Mori H., Yura T., Ishihama A.;
RT      "Systematic sequencing of the Escherichia coli genome: analysis of
RL      the 2.4-4.1 mln (110,917-193,643 bp) region.";
RN      Nucleic Acids Res. 22:1637-1639(1994).
RN      [2]
RP      SEQUENCE FROM N.A.
RC      STRAIN=K12 / MG1655;
RX      MEDLINE=97426617; PubMed=9278503;
RA      Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RT      Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA      Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RT      Mau B., Shao Y.;
RL      "The complete genome sequence of Escherichia coli K-12.";
SC      Science 277:1453-1474(1997).
CC      -1. SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. STRONG, TO
CC      RHIZOBIACEA NODI.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL collaboration -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb.ch/announce/
CC      or send an email to license@isb.ch).
CC      -----
CC      EMBL; D26562; BAA05583.1; -
DR      EMBL; AE000122; AAC73238.1; -
DR      PIR; S45204; S45204.
DR      EcoGene; EG12320; yadG.
DR      InterPro; IPR003593; AAA_ATPase.
DR      InterPro; IPR003439; ABC_transport.
DR      Pfam; PF00005; ABC_tran; 1.
DR      ProDom; PD000006; ABC_transport; 1.
DR      SMART; SM00382; AAA; 1.
DR      PROSITE; PS00211; ABC_TRANSPORTER; 1.
KW      Hypothetical protein; ATP-binding; Transport; Complete proteome.
FT      NP_BIND 38 45 ATP (POTENTIAL).
SO      SEQUENCE 308 AA; 34647 MW; 9EF43C2BFBFF3E47 CRC64;

Query Match 2.98; Score 347; DB 1; Length 308;
Best Local Similarity 28.28; Pred. No. 1.9e-14;
Matches 104; Conservative 64; Mismatches 133; Indels 68; Gaps 8

OY      895 LKIGVSIQNLVKYVRDGGKVAVDGIALNFEQGQTTSFLGHNGAGKTTMSILNGLPPTS 954
DB      1 MTALELDQLKLTYPGQVQ-ALRQIDQLQVEAGDEYALLGPAGACKSYTIGISSLVAKTS 59
OY      955 GTAYVLGKDIRSEMSTIRQNTGVCPOHNVLPDMLTVEHITWFAVRLKGLSK--HYVAM 1012
DB      60 GRVSYFVGDLDEVDLVNARROGLVPOEFNPFETVOOILVNOGYGVGEKKEVITSEK 119
OY      1013 EQMALDVGIPSSKLSKTSKTSOLSGQMOKLSVALAFVGGSKVVLDEPTAGVDPYSRGIW 1072
DB      120 YLKQIDLW---CKRNRARMLSGMKRRMLIARALMHPKLLILDEPTAGVDIELRSMW 176
OY      1073 ELLLYR-QGRTIILSTHHMEDADVLDGRILAISHGKLCCVGSISFLKNOJGTGYLLTV 1131
DB      177 GLKLDLNKGTIILITLTLYLEAEAMLCNIGIGIQGEVY- 215
OY      1132 KKDVESSLSCNNSSTVSYLKEKEDVSQSSDAGLSDHESDYLITDVAISMLIRKHV 1191
DB      216 -----ENTSMKALLAKIKSEFTILDAPKSPILKIDG 247

```

```

QY      1192 SEATTLEIGHETLYVLPPEAKEGAFVELFPHIDRSLDGLS--SYGISTTLEELFL 1249
Db      248 YQYLAVD-----TATLEVEVLREOGINSVF-----TOLSEOGIQLVMRNKRNLBELFLV 297
OY      1250 KVAEEGVGD 1258
Db      298 SLVNEKGGD 306

RESULT 12
NODL_RHILLO
ID NODL_RHILLO STANDARD; PRT; 340 AA.
AC NP_23703.1
DT 01-NOV-1991 (Rel. 20, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Modulation ATP-binding protein I.
GN NODI OR MLR6164.
OS Rhizobium lotii (Mesorhizobium lotii).
OC Bacteria; Proteobacteriae; alpha subdivision; Rhizobiaceae group;
OC Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxID=381;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NZP 2213;
RA MEDLINE=91067466; PubMed=2251131; Terzaghi E.A., Scott D.B.;
RA Young C.A., Collins-Emerison J.M., Tetzaghi E.A., Scott D.B.;
RT "Nucleotide sequence of Rhizobium lotii nodI."
RL Nucleic Acids Res. 18:6691-6691(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF30309;
RA MEDLINE=21082930; PubMed=11214968;
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimojo S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT Mesorhizobium lotii.";
RL DNA Res. 7:331-338(2000).
CC -!- FUNCTION: FORMS, WITH NODJ, A MEMBRANE TRANSPORT COMPLEX INVOLVED
CC IN THE MODULATION PROCESS. IT PROBABLY EXPORT A MODIFIED BETA-1,4-
CC LINKED N-ACETYLGALACTOSAMINE OLIGOSACCHARIDE.
CC -!- SUBCELLULAR LOCATION: Membrane-associated
CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
-----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outpost at
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC CC EMBL: X55705; GAA39236.1; ALU_INIT.
DR DR EMBL: AP003008; BAB52501.1; -.
DR DR PIR: S13590; S13590.
DR DR HSSP: Q58663; IG6H.
DR DR InterPro: IPR003593; AAA_ATPase.
DR DR InterPro: IPR003439; ABC_transportr.
DR DR Pfam: PF00005; ABC_tran.1.
DR DR ProDom: PD000006; ABC_transportr.1.
DR DR SMART: SMO0382; AAA.1.
DR DR TIGRFAMS: TIGR01289; nodI.1.
DR DR PROSITE: PS00211; ABC_TRANSPORTER.1.
KW KW Modulation; Membrane; ATP-binding; Transport; Complete proteome.
FT FT NP_BIND 74 81 ATP (BY SIMILARITY).
FT FT CONFLICT 10 10 D->E (IN REF. 1).
FT FT CONFLICT 97 97 T->A (IN REF. 1).
FT FT CONFLICT 129 129 F->L (IN REF. 1).
FT FT CONFLICT 167 167 D->N (IN REF. 1).
SQ SEQUENCE 340 AA: 37428 MW: 57777722B8D130EA CRC64:

```

Query Match 2.9%; Score 344.5; DB 1; Length 340;
 Best Local Similarity 32.9%; Pred. No. 3.2e-14;
 Matches 80; Conservative 51; Mismatches 93; Indels 19; Gaps 4;

OY 867 GSESEKHPGSGNKRMSKMECEEPHLKLVSIQMLVYRGDMVAVDGLALNPFYGS 926
 DB 25 GOTSASSVPSAS-----TVAVDPAVTSY--GKIVDELSTFSVASG 67
 OY 927 QITSPFGNAGKTTTMSILTLGTFPTSGTAYILGKDIKSEKSTINOMLGVCPQHNVLDP 986
 DB 68 ECFGLGNGAGKSTIARMLGMCPPAGITVLGVVPARARLARRGIGVQPDNDQ 127
 OY 987 MLTYEHIWYFARLKLSEKHVKAEMOMALDVLSPSSKLKSTQSOLSGMQRKLSVALA 1046
 DB 128 EFTYRENLVGFYFGKSTQSEAVIPSL-LEFARLERKADAVSELGSGMKRCLTMARA 186
 OY 1047 FVGSKVVIIIDEPYAGDPSRKGIME-LLKTYRGRTITLLSTHMDADVGLDRIATIS 1105
 DB 187 LINDPOLIVDEPTTGLDPAHRLIWERLRLALLARGKITIILTHFWEAEKLCDFLCVLE 246
 OY 1106 HGR 1108
 DB 247 KGR 249

RESULT 13
 MODL_RHIGA STANDARD; PRT: 347 AA.

AC P50332;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Modulation ATP-binding protein I.
 GN NOD1.
 OS Rhizobium galegae.
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Rhizobiaceae; Rhizobium.
 OX NCBI_TaxID=399;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=HAMBI 1174;
 RA MEDLINE=99403395; PubMed=10474187;
 RA Suominen L., Paulin L., Saano A., Saren A.M., Tas E., Lindstrom K.;
 RT "Identification of nodulation promoter (nod-box) regions of Rhizobium
 RT galegae.";
 RT FEBS Microbiol. Lett. 177:217-223(1999).
 CC -I- FUNCTION: FORMS, WITH NOD1, A MEMBRANE TRANSPORT COMPLEX INVOLVED
 CC IN THE NODULATION PROCESS. IT PROBABLY EXPORT A MODIFIED BETA-1,4-
 CC LINKED N-ACETYLGALACTOSAMINE OLIGOSACCHARIDE.
 CC -I- SUBCELLULAR LOCATION: Membrane-associated.
 CC -I- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement. (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X87578; CA60881.1;
 DR InterPro: IPR003593; AAA_ATPase.
 DR InterPro: IPR003439; ABC_Transporter.
 DR Pfam: PF00005; ABC_tran; 1.
 DR ProDom: PD000006; ABC_transportr; 1.
 DR SMART: SM00382; AAA; 1.
 DR TIGRfam: tigr01288; AAA; 1.
 DR PROSITE: PS00211; ABC_TRANSPORTER; 1.
 KM Modulation: Membrane; ATP-binding; Transport.
 FT NE_BIND 81 ATP (POTENTIAL).
 SO SEQUENCE 347 AA; 38435 MW; AC91210C44C9A6C6C64;

Query Match 2.9%; Score 343.5; DB 1; Length 347;
 Best Local Similarity 30.0%; Pred. No. 3.8e-14;
 Matches 86; Conservative 58; Mismatches 126; Indels 17; Gaps 6;

OY 1873 REFIRPRPV---NAKLSPLNDEDEDVARRRQRIILDDGGQNDILEIKELKTYRRKRPAY 1929
 DB 7 REMLRPKTIAMDQASASARSNPREIKTIGLLEPASASAPPAIDQAVMYR--DKTVV 64
 OY 1930 DRIVCPPECCGGLGVNAGKSGSTFKMTGPTVTRGDAPLNNKLSLSNIEHWONG 1989
 DB 65 DLSFGRACECGGLGPNAGKSTITRMILGWATPSACKISVLGIPVGKARLARASIG 124
 OY 1990 YCPQDAITLLTGRHV---EFALLRGVEPEKGVKGEAATIKLGVKGEYAGNY 2045
 DB 125 VVSQFDLMEFVFRNLNLFVGYFQMSRALEKLPISLLEFA---OLEAKADRVSDL 180
 OY 2046 SGNKRRSLTAMALIGPPVFLDEPTTGMDPKARFLMNCALSVYKGRSVLTSHME 2105
 DB 181 SGKRRRLTLARALVNDPOLLILDEPTTGLDPPARHQIWERLRLIRGKITIILTHMD 240
 OY 2106 ECEALCTRMAMVNGRFRCLG--SVQHLKRRPDGTTIVRIAGSNPD 2151
 DB 241 EAERMCDRLCVLEGGMIAEGPPLSLIEDITG---CPVIEVYGGNPD 284

RESULT 14
 MODL_HUMAN STANDARD; PRT: 1280 AA.

AC P08183; Q12755; Q14812;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Multidrug resistance protein 1 (P-glycoprotein 1) (CD243 antigen).
 GN ABCB1 OR PGLY1 OR MDR1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=87028230; PubMed=2876781;
 RA Chen C.-J., Chin J.E., Ueda K., Clark D.P., Pastan I., Gottesman M.M.,
 RA Roninson I.B.;
 RT "Internal duplication and homology with bacterial transport proteins
 RT in the mdr1 (P-glycoprotein) gene from multidrug-resistant human
 RT cells.";
 RT Cell 47:381-389(1986).
 RL [2]
 RN SEQUENCE FROM N.A.
 RP MEDLINE=90094448; PubMed=1967175;
 RA Chen C.-J., Clark D.P., Ueda K., Pastan I., Gottesman M.M.,
 RA Roninson I.B.;
 RT "Genomic organization of the human multidrug resistance (MDR1) gene
 RT and origin of P-glycoproteins.";
 RT J. Biol. Chem. 265:506-514(1990).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=97190336; PubMed=9038218;
 RA Chen G., Duran G.E., Steger K.A., Lacayo N.J., Jaffrezou J.P.,
 RA Dumontet C., Sikic B.I.;
 RT "Multidrug-resistant human sarcoma cells with a mutant P-glycoprotein,
 RT altered phenotype, and resistance to cyclosporins.";
 RT J. Biol. Chem. 272:5974-5982(1997).
 RN [4]
 RP SEQUENCE OF 1-234 FROM N.A.
 RA Smith A., Beck C., Gibson A.;
 RL Submitted (Aug-1997) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE OF 178-215 AND 800-856 FROM N.A.
 RX MEDLINE=90290529; PubMed=1972623;
 RA Gekeler V., Weger S., Probst H.;
 RT "mdr1/P-glycoprotein gene segments analyzed from various human
 RT leukemic cell lines exhibiting different multidrug resistance


```

QY 1212 AAKGAFVLEFHEIDRLSDIGISSYGISETTLEIFLKVAESGVDAEFTSDGTLFARRN 1271
Db 666 -----
QY 1272 RAACDCKSCLRPTEDDAADPNDSIDIPESRETDLSGMDGKGYOVKMKLTQOOFA 1331
Db 669 RRSVGSQAORRKTSTKEAL---DESLPP-----VSWMRLKMLNTE 707
QY 1332 LMKRLLIARRSRKGFQAIYLAIVFACIALVSLVPPPGKYPSLELOFMVNEOYTFV 1391
Db 708 --WPEFV-----GVFCALINGLOPAFAAIIFSKII----- 736
QY 1392 SNADEDTGTELLNALTKPGEGFCROMEGNPIDPFCQAGEPEWMTAPVPTIMDLPON 1451
Db 737 -----GVFTR----- 741
QY 1452 GNMWTMNPSPACQSSSKIKMLPVCPCGAGLPPROKONTADILQDLTGRIISDYLVK 1511
Db 742 ----- 762
QY 1512 TYVOLIASKIKNKIWNFEFYGSGSLGVSNTOALPSSQVNDAIKMKHKLKLAADS-A 1570
Db 763 GLISFII-----FLOGFTFGKA-----GELITRLRYMWFERSMLRODVSMF 804
QY 1571 DRFLNSIGREMTGLDRNNKVKVFNKGMHAISFLVNNALILRANLOKGENPSHYGT 1630
Db 805 DDPKNTGALITRL-ANDAQV---KG--AIGSRLAVITQNI--ANLGTI-----II 849
QY 1631 AFNHPMLNLTQOQLEVALMTSDVLSVCIFFAMSPFASVVLIDERSKAKHLOFT 1690
Db 850 SFYIGMQLT-----LLILAVPLIAGVAK-----MLSGQALKKKELE-- 891
QY 1691 SGKPIYIWNLSNFWDMCVNVPALVILIFICFOOK---SVYSSNIPVIALI--LLEY 1745
Db 892 GACKIATEAIENF-----RIVVSLTQKFEHMYAQSLOVFPYNSLRKAHIF 938
QY 1746 G--WSTIPLMYPAFVEFKIPSTAYVVLVSVNLFIGINSVATFVLEFTDNKLNINDIL 1803
Db 939 GIFSFSTQAMMYFSAAGCFRFAIYLA-----HKLSFEDVL 975
QY 1804 KSVFLLEFPHFCIGRLIDMVKNOAMADALERGENRFV---SPLSMDLVGRMLFAMAV 1858
Db 976 ---LVESAIVGAMAVGVQVSSFADPYAKAKISAHIIMIIEKPLIDSYSTEGMLMNTL 1931
QY 1859 EGVEFELLVLIOYRFFRPR-PIVAKISPLDEDEVDREKORILDGCGONDILEIKEL 1917
Db 1032 EGNVTFGEVY---FNYTPRPDIPLYGLS-----LEVKK- 1062
QY 1918 TTIYRRKKRPVADRICVGIPEGEFGLGVNAGKSSIFKMLTGDTTVTRGDAFLKNSI 1977
Db 1063 -----GOTLALVSSGCGKSTVYOLLERFYPLAGKYLLOKET 1101
QY 1978 LS-NIHEVHOMNGYCPQ---FD-----ATTELLIGREHVEFFALLRGV 2016
Db 1102 KRLNQMILRAHIGVSOEPILIFDCSIAENIAYGDSRVSQSEIFVRAKBAHIAHFESE 1161
QY 2017 PEKEVKGVEMAIRKLGIVKGEKYAGNYSGGNKRKLTAMALIGGPVYVLEDEPTGMD 2076
Db 1162 PNKYSTKVGDKGTO-----LSGQKORAIKALVROPHILLIDEATSAID 1207
QY 2077 PKARRFLMNCALSVKGRSVVLTSHECEBALCTRMALMVNGRRPLGSGVOHLNRRG 2136
Db 1208 TESKEVQOE-ALDKARGEFCIVIAHRLSTION-ADLIVFONGRVKENGTHQOULLAQNG 1265
QY 2137 DGYTIVIRIAGS 2148
Db 1266 IYFSMVSVOAGT 1277

```

```

RESULT 15
NODI_RHIME
ID NODI_RHIME STANDARD; PRT: 355 AA.
AC 052618;

```

```

DT 30-MAY-2000 (Rel. 39, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Nucleotide sequence of nodJ region of Rhizobium meliloti pSymA.
GN NODI OR RA0472 OR SMA0864.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OG Plasmid pSymA (megaplasmid 1).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21396509; PubMed=11481432;
RA Barnett M.J., Fisher R.F., Jones T., Komp C., Abola A.P.,
RA Barley-Hubler F., Bowser L., Capela D., Galibert F., Gouzy J.,
RA Gurjal M., Hong A., Huzar L., Hyman R.W., Kahn D., Kahn M.L.,
RA Kaiman S., Keating D.H., Palm C., Peck M.C., Surzycki R., Wells D.H.,
RA Yeh K.-C., Davis R.W., Federspiel N.A., Long S.R.;
RT "Nucleotide sequence and predicted functions of the entire
RT Sinorhizobium meliloti pSymA megaplasmid."
RL Proc. Natl. Acad. Sci. U.S.A. 98:9883-9888(2001).
RN [2]
RP SEQUENCE OF 143-355 FROM N.A.
RC STRAIN=1021;
RA Barnett M.J., Long S.R.;
RT "Nucleotide sequence of nodJ region of Rhizobium meliloti pSymA."
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: FORMS, WITH NODJ, A MEMBRANE TRANSPORT COMPLEX INVOLVED
CC IN THE NODULATION PROCESS. IT PROBABLY EXPORT A MODIFIED BETA-1,4-
CC LINKED N-ACETYLGUCCOSAMINE OLIGOSACCHARIDE.
CC -1- SUBCELLULAR LOCATION: Membrane-associated.
CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; AE007237; AAK65130.1; -
CC DR EMBL; AF043118; AAB97762.1; -
CC DR InterPro: IPR003593; AAA_ATPase.
CC DR InterPro: IPR003439; ABC_transportr.
CC DR Pfam; PF00005; ABC_tran; 1.
CC DR ProDom; PD000006; ABC_transportr; 1.
CC DR SMART; SM00382; AAA; 1.
CC DR TIGRFAMs; TIGR01288; nodJ; 1.
CC DR PROSITE; PS00211; ABC_TRANSPORTER; 1.
CC DR Nucleotide; Membrane; ATP-binding; Transport; Plasmid;
CC KW Complete proteome.
CC FT NP_BIND 89 ATP (POTENTIAL).
CC SEQUENCE 355 AA; 39268 MW; 4DC869D98C35DC CRC64;

```

```

Query Match 2.8%; Score 331.5; DB 1; Length 355;
Best Local Similarity 33.9%; Pred. No. 2.3e-13;
Matches 74; Conservative 52; Mismatches 85; Indels 7; Gaps 4;

```

```

QY 891 EPHLKIGVSIQMLVKYRDGMKVAADGALNFEQOITSFGHNGAGTMTSTITGLF 950
Db 52 KPT---VAIDVASVTKSYGD--KPVINGLSFIVAAAGECGGLGPGACKSTITRAIILGMT 106
QY 951 PPTSGTAYIIGKDIRSEMSITRONIGVCQOHVLDMLTVEHIFVYRLKLGSEKHVKA 1010
Db 107 TPGTGELTYVGVPPSRKARLRARIGVVPQFNLDLEFVRENLLVFRYRPMSTREIEA 166
QY 1011 EMDQALDVGLPSSKLKSKTSQSLSGGMQRKLSVALAFVGGSKVVLIDPPTGVPYSRG 1070
Db 167 VIPSL-LEFARLENKADARVSDLSGGMKRRLTLARALINDPOLLLDEPTTGLDPHAHL 225
QY 1071 IWEELLK-YRQGTILLSTHMDADVLDGRIAIISHG 1107

```

Fri Mar 7 09:16:54 2003

us-09-595-526c-2.rsp

Page 21

```

    ||| | . :|||:||| | ||: :: |
Db   226 IWEKRLSLRGKTILLTHIMEARLCLDRCLVLEAG 263

Search completed: March 7, 2003, 08:48:05
Job time : 61 secs
```


GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 7, 2003, 08:44:14 ; Search time 64 Seconds
(without alignments)
3396.249 Million cell updates/sec

Title: US-09-595-526c-2

Perfect score: 11797
Sequence: 1 MACWPOLRLMLKMLTFRRR.....VDVAVLTSFLQDEKVESTY 2261

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 segs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database:

PIR-73: *
1: PIR1: *
2: PIR2: *
3: PIR3: *
4: PIR4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	10922	92.6	2201	2	A54774	ATP binding cassette
2	3345.5	28.4	1529	2	A59189	ATP-binding cassette
3	3129.5	26.5	1472	2	B54774	ATP binding cassette
4	2645.5	22.4	1704	2	S71363	probable ATP-bind
5	2642.5	22.4	1704	2	A59188	ATP-binding cassette
6	2060	17.5	1802	2	T33783	hypothetical prote
7	1984	16.8	1816	2	A84845	probable ABC trans
8	1863.5	15.8	1447	2	T15200	hypothetical prote
9	1792	15.2	1317	2	C88925	protein F33E1.4 [
10	1542	13.1	1758	2	F88559	ATP-binding cassel
11	1538.5	13.0	1704	2	T42749	transport protein
12	1538.5	13.0	1767	2	S60124	hypothetical prote
13	1402	11.9	1246	2	T00826	hypothetical prote
14	1158.5	8.6	1564	2	T27121	hypothetical prote
15	1013.5	8.6	373	2	T42750	hypothetical prote
16	1010.5	8.6	1431	2	T42748	hypothetical prote
17	846	7.2	269	2	T46467	hypothetical prote
18	788.5	6.7	1011	2	T07712	probable ABC-type
19	693	5.9	900	2	T07717	probable ABC-type
20	667	5.7	895	2	T07714	probable ABC-type
21	656.5	5.6	722	2	T07716	probable ABC-type
22	645	5.5	925	2	T07713	probable ABC-type
23	624.5	5.3	1336	2	T18288	hypothetical prote
24	485	4.1	196	2	T12512	hypothetical prote
25	432	3.7	339	2	S74048	probable daonorubi
26	428	3.6	664	2	T07715	probable ABC-type
27	426.5	3.6	324	2	C71081	probable resistanc
28	421	3.5	328	2	E75108	daonorubcin resis
29	411	3.5	330	2	S27707	daonorubcin resis

30	410.5	3.5	310	2	C96929	ABC-type multiding
31	410.5	3.5	327	2	D72257	hypothetical prote
32	410	3.5	314	2	D97318	ABC-type MDR trans
33	409.5	3.5	398	2	C69485	daonorubcin resis
34	403.5	3.4	297	2	AE1816	ABC transporter (A
35	396.5	3.4	333	2	D72492	probable ABC trans
36	395.5	3.4	246	2	S75436	hypothetical prote
37	390	3.3	301	2	E72384	ABC transporter, A
38	388.5	3.3	310	2	E69920	ABC transporter (A
39	388	3.3	311	2	G69803	ABC-type ABC-type
40	386.5	3.3	727	2	T07718	probable MDR trans
41	386	3.3	259	2	H97301	ABC-type MDR trans
42	386	3.3	331	2	D70984	probable dtra prot
43	385	3.3	350	2	B69065	ABC transporter (A
44	384	3.3	331	2	S73019	daonorubcin resis
45	380	3.2	312	2	H97342	ABC-type MDR trans

ALIGNMENTS

RESULT 1

A54774
ATP binding cassette transporter ABC1 - mouse
C:Species: Mus musculus (house mouse)
C:Date: 05-Apr-1995 #sequence_revision 05-Apr-1995 #text_change 02-Feb-2001
C:Accession: A54774
R:Luciani, M.F.; Denizot, F.; Savary, S.; Mattei, M.G.; Chimini, G.
Genomics 21, 150-159, 1994
A:Title: Cloning of two novel ABC transporters mapping on human chromosome 9.
A:Reference number: A54774; M01D:94375008; PMID:8088782
A:Accession: A54774
A:Molecule type: mRNA
A:Residues: 1-2201 <LUC>
A:Cross-references: GB:X75926; NID:9495256; PIDN:CAA53530.1; PID:9495257
C:Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homolog
C:Keywords: ATP; duplication; nucleotide binding; P-loop
F:856-1047/Domain: ATP-binding cassette protein <ABC1>
F:873-880/Region: nucleotide-binding motif A (P-loop)
F:1869-2060/Domain: ATP-binding cassette homology <ABC2>
F:1886-1893/Region: nucleotide-binding motif A (P-loop)

Query Match	Score	Length	DB 2:	Length	DB 1:
Best Local Similarity	94.9%	Pred. No. 0:			
Matches 2089; Conservative 55; Mismatches 57; Indels 0; Gaps 0;					
QY 61	MPAGTLPPWVGITCNANPCFRYPGGAAGVGNFNKSIYARLFSARLLYSQDT	120			
DB 1	MPAGTLPPWVGITCNANPCFRYPGGAAGVGNFNKSIYARLFSARLLYSQDT	60			
QY 121	SMKDKRYLRTLOQIKKSSNKLQDLVNEFFSGFLYHNLSLPKSTYDKMLRADVTLH	180			
DB	SIKDKRYLRTLOQIKKSSNKLQDLVNEFFSGFLYHNLSLPKSTYDKMLRADVTLH	120			
QY 181	KVFLQGYDLHLTSLCNKSSSEMTOLGDOVESELGIPKREKLAARVLRNSMDILKPLT	240			
DB	KVFLQGYDLHLTSLCNKSSSEMTOLGDOVESELGIPKREKLAARVLRNSMDILKPLT	180			
QY 241	RTLNSTSPFPKELAEAVKTLHSLGLTADLFPSKRSWSMRQEWMLFTVNVSSSSSTOI	300			
DB	RTLNSTSHLPQHLAEATVTLVLDLSGLADLFSTKSMRQEWMLFTVNVSSSSSTOI	240			
QY 301	YOAVSRIVCGHREGGGLIKSLNMYEDNNYKALFGNGTEDEATFPYDNTSTPYCDLMLK	360			
DB	YOAVSRIVCGHREGGGLIKSLNMYEDNNYKALFGNGTEDEATFPYDNTSTPYCDLMLK	300			
QY 361	NESSPSLRILKALKPLLVGKILYTPDTPATROYMAEVNKTFOELAVFDLQGMWELS	420			
DB	NESSPSLRILKALKPLLVGKILYTPDTPATROYMAEVNKTFOELAVFDLQGMWELS	360			
QY 421	PKIWTFMENSQEMDLVRLMLDSRDNDHFWDQDLGMDTADIVAFIAKAPDEVQSSNGS	480			
DB	PKIWTFMENSQEMDLVRLMLDSRDNDHFWDQDLGMDTADIVAFIAKAPDEVQSSNGS	361			

QY	481	VTYREAFENENOAIRISFMECVMLNKLEPLATEWLLINKSMELDERKFMAGIVFTG	540
Db	421	VYTRKEAFNETNOALQIISRFMECVMLNKLEPIPLEVRLINKSMELDERKFMAGIVFTG	480
QY	541	ITPGSIELPHHYVKKIRIMDIDNVERNRNKIKDQWMPDGPADPEEDMRVWKGFAVLDDV	600
Db	481	ITPDSVELPHHYVKKIRIMDIDNVERNRNKIKDQWMPDGPADPEEDMRVWKGFAVLDDV	540
QY	601	EOALIRVLCTEKKTKGYVMOQMPYCYVDJFLRWSSSMPLEMTLAMIYSAVATIGIV	660
Db	541	EOALIRVLCTEKKTKGYVMOQMPYCYVDJFLRWSSSMPLEMTLAMIYSAVATIGIV	600
QY	661	YEKEARLEKTMIRIMGDINSILMFWSFISLPIELVASAGLVLVILKGLNLPYSDPSVVEV	720
Db	601	YEKEARLEKTMIRIMGDINSILMFWSFISLPIELVASAGLVLVILKGLNLPYSDPSVVEV	660
QY	721	FLSVFAVVTILQCFLLISTLFSRANILAAACGIIYFTLIPVLCVAMQDVGFEFLKIFAS	780
Db	661	FLSVFAVVTILQCFLLISTLFSRANILAAACGIIYFTLIPVLCVAMQDVGFEFLKIFAS	720
QY	781	LILSPAFAFGCEFALEFEOGIGVOWDMLFSPVPEOESFNLTJISMLPDTLYGVMW	840
Db	721	LILSPAFAFGCEFALEFEOGIGVOWDMLFSPVPEOESFNLTJISMLPDTLYGVMW	780
QY	841	YITAVFPGQYGIIPRWYFPCTKSIFYWGESEDESKSHPSNOKRMSEIIMEEPHLKIGVS	900
Db	781	YITAVFPGQYGIIPRWYFPCTKSIFYWGESEDESKSHPSNOKRMSEIIMEEPHLKIGVS	840
QY	901	IONLVYVYEDGKAVVADGIALNFVEGQTTSLFLNGNGAKTTMSILTLGLFPPTSGAYTL	960
Db	841	IONLVYVYEDGKAVVADGIALNFVEGQTTSLFLNGNGAKTTMSILTLGLFPPTSGAYTL	900
QY	961	GKDISEMSTIRONLGYCPOHNVLFDMLTYEHEIMFAPARKIGISEKHVYAEKMOALDVG	1020
Db	901	GKDISEMSTIRONLGYCPOHNVLFDMLTYEHEIMFAPARKIGISEKHVYAEKMOALDVG	960
QY	1021	LPSSTLKSTLSOGSGMORKLSVALAVYSGSKYVILDEPTIYADVPSRRGIMELLKTRYQ	1080
Db	961	LPSSTLKSTLSOGSGMORKLSVALAVYSGSKYVILDEPTIYADVPSRRGIMELLKTRYQ	1020
QY	1081	GRTTILSTHMHDEADVLCDRLAIISHEKLCVCSGLEKMOIGTGYVILTYLKKDVESSIS	1140
Db	1021	GRTTILSTHMHDEADVLCDRLAIISHEKLCVCSGLEKMOIGTGYVILTYLKKDVESSIS	1080
QY	1141	SCRNSSSTVYLTKKEDSVSSSAGISGSHESDRTLIDVSAISNLRKHSSEARLVEDI	1200
Db	1081	SCRNSSSTVYLTKKEDSVSSSAGISGSHESDRTLIDVSAISNLRKHSSEARLVEDI	1140
QY	1201	GHELTLYVLYEAAKEGAFVELFHEIDRLDLDISSYGISSEITLLEIFLKVAAESGVAE	1260
Db	1141	GHELTLYVLYEAAKEGAFVELFHEIDRLDLDISSYGISSEITLLEIFLKVAAESGVAE	1200
QY	1261	TSDDGLPARRRRRRRFGKQKOCLEPFFEDDAADPNOSDIDPESRETDLLSGMDKOSGYVK	1320
Db	1201	TSDDGLPARRRRRRRFGKQKOCLEPFFEDDAADPNOSDIDPESRETDLLSGMDKOSGYVK	1260
QY	1321	GMKLTQOOFVALLMKRLLIARRSRKCFPAOIVLPYAFVCIATVFSLIYPPGKTPSLEQ	1380
Db	1261	GMKLTQOOFVALLMKRLLIARRSRKCFPAOIVLPYAFVCIATVFSLIYPPGKTPSLEQ	1320
QY	1381	PWMYNEOYTVSNDAPEDDTGTELELNAITLKDPGFSTRCEMGNPJPDPFQCAQGEEMTAP	1440
Db	1321	PWMYNEOYTVSNDAPEDDTGTELELNAITLKDPGFSTRCEMGNPJPDPFQCAQGEEMTAP	1380
QY	1441	VPOGITMLFONGMWMTOMNSPACQCSSDIKIKMLPVCPPGAGGLPPOROKONTADILQDL	1500
Db	1381	VPOGITMLFONGMWMTOMNSPACQCSSDIKIKMLPVCPPGAGGLPPOROKONTADILQDL	1440
QY	1501	TGRNISDIYLVKTYVOILIAASLANKIMVNEFRYGGFSLVSNVMOALPPOSEFINAIKQMK	1560
Db	1441	TGRNISDIYLVKTYVOILIAASLANKIMVNEFRYGGFSLVSNVMOALPPOSEFINAIKQMK	1500

Qy	1561	HELEADSDPEPLSLICREMTGLDTRNNKXVWFNNKGHAISSFLNVLNNAITLRANLQK	1620
Dp	1501	LKLTJDTSDADPELSSJGCPMAGJIDTCKNNKXKWFENNGHAISSFLNVLNNAITLRANLQK	1560
Qy	1621	GENPSHYGTAEENHNLNLTJKOOLSEVALAMTTSYDVLSICVIFAMSFVPASVVFJDER	1680
Dp	1561	GENPSSOYGITTAENHNLNLTJKOOLSEVALMTTSYDVLSICVIFAMSFVPASVVFJDER	1620
Qy	1681	VSKAKHLOPISGVKRVITWLSNFWDMCNVVPATVLIITFICFOOKSVYSTNLPVIAL	1740
Dp	1621	VSKAKHLOPISGVKRVITWLSNFWDMCNVVPATVLIITFICFOOKSVYSTNLPVIAL	1680
Qy	1741	LLLLGWSITPLMPYPASVFEKIPSTAVVLTSNLFGINGSVAPFVLELFTDNLNIN	1800
Dp	1681	LLLLGWSITPLMPYPASVFEKIPSTAVVLTSNLFGINGSVAPFVLELFTNNLNIN	1740
Qy	1801	DLTKSVFLIFPHFCLGRLIDMKNQAMADALERFGENRFSPLSMDLVRNLFMAVEG	1860
Dp	1741	DLTKSVFLIFPHFCLGRLIDMKNQAMADALERFGENRFSPLSMDLVRNLFMAVEG	1800
Qy	1861	VVEFLITVLIDYREFLIPRPVNAKLSPLNDEDVDVAREKORILLDGGONDILELKELT	1920
Dp	1801	VVEFLITVLIDYREFLIPRPVNAKLSPLNDEDVDVAREKORILLDGGONDILELKELT	1860
Qy	1921	YRRKKRPVNDICVGIIPGECFGLGVNAGKSSTFKMLTGDITVTRGDAFLNNNSILSN	1980
Dp	1861	YRRKKRPVNDICVGIIPGECFGLGVNAGKSSTFKMLTGDITVTRGDAFLNNNSILSN	1920
Qy	1981	IHEYHONNGYPOPDATTELLTGREHVEFPALLRGVPEKVEYKVEEAMIRKLGIYKGEK	2040
Dp	1921	IHEYHONNGYPOPDATTELLTGREHVEFPALLRGVPEKVEYKVEEAMIRKLGIYKGEK	1980
Qy	2041	YAGVSGGNKRRKISTAMALIGPPVLEDPPTGMDPKARFRFLMNCALSYVKEGRSVLT	2100
Dp	1981	YASVSGGNKRRKISTAMALIGGPPVLEDPPTGMDPKARFRFLMNCALSIYKEGRSVLT	2040
Qy	2101	SHSMEEEALCTNRALIMVNGRFKJGSYOHKKNRGDGYTTIVRLAASNPLDKPVODFG	2160
Dp	2041	SHSMEEEALCTNRALIMVNGRFKJGSYOHKKNRGDGYTTIVRLAASNPLDKPVODFG	2100
Qy	2161	LAFGVSYLAKERHNNLLOYLPSLISLARIFSLISQSKRLHIDYVSOTTTIDQVFN	2220
Dp	2101	LAFGVSYLAKERHNNLLOYLPSLISLARIFSLISQSKRLHIDYVSOTTTIDQVFN	2160
Qy	2221	AKDSDDDHLKDLSTLHKNOTVVDVAVLTSFLQDEKVEKSY	2261
Dp	2161	AKDSDDDHLKDLSTLHKNOTVVDVAVLTSFLQDEKVEKSY	2201

```

RESULT 2
A59189
ATP-binding cassette transporter - human (fragment)
N/Alternate names: KIAA1062; protein
C/Species: Homo sapiens (man)
C/Date: 18-Feb-2000 #sequence_revision 18-Feb-2000 #ext_change 02-Jun-2000
C/Accession: A59189
R/Rikunou, R.; Megase, T.; Ishikawa, K.; Hirotsawa, M.; Miyajima, N.; Tanaka, A.; Kotan
DNA Res. 6, 197-205, 1999
A/Title: Prediction of the coding sequences of unidentified human genes. XIV. The com
A/Reference number: Z22961; MUID:99397452; PMID:10470851
A/Accession: A59189
A/status: preliminary; not compared with conceptual translation
A:molecule type: mRNA
A:Residues: 1-1529 <KTK>
A:Cross-references: GB:A0028985; MID:g55689460; PIDN:BAAB3014.1; PID:d1046841; PID:g56
A:Experimental source: chromosome 9; clone hJ03579; clone 11b pBluescriptII SK plus;
C/genetics:
A:Map position: 9
A:Note: KIAA1062
C:Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homolog
Query Match 28.4%; Score 3345.5; DB 2; Length 1529;
Best Local Similarity 46.7%; Pred. No.4.1e-208;

```

Matches 721; Conservative 207; Mismatches 381; Indels 235; Gaps 33;

```

QY 836 GWTWYIENAFVGGYQYGRPMWYFPCRTKSYMFE---ESDEKSHFGSNOKRMS-----ETC 887
   1:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 GLTWTIEAVHNGMYGLDRPMWYFPLQKSYWLGSGRTKEMEMSWMARTPRLSWMEEDQAC 60
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 888 -----MEEPHKLKGVSIQNLVYVRDGMKVAVDGALNLFEGQITSFLGHN 935
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 61 AMESRREETROMEEPRHPLPLVVCVDKLTLYKKDKKLALNKLKSLINLENQVVSFLGHN 120
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 936 GACKTTMSILGGLPPTSGTAYILGKDIRSEMTIRONLGVCPQHNVLFPMLVEEHIM 995
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 121 GACKTTMSILGGLPPTSGSATIYGHDIRTMDLRKLGMCPOHNVLFPRLTYVEEHLIM 180
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 996 FYARLGLSEKHVAKMEOMALDVGLPSSKLSKTSQSLSGGMORLSTVALAFVGSKVYI 1055
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 181 FYSRLSKMAQEEIRREMDKIEDLEL-SNKRHSLVOTLSGKKRKLSTVALAFVGSRAI 239
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1056 LDEPTAGVDYSRGIMWELLKRYRQRTIILSTHMDADVLGDRITAIISHGKLCVSS 1115
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 240 LDEPTAGVDYARAIIMDLTKKPGRTIILSTHMDADVLGDRITAIISHGKLCVSS 299
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1116 LFLKNOLGTGYTLVKKDVESLSCRNSSIVSTLKEDVSQSSDAGLGSDESDT 1175
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 300 LFLKGTGDDGRLTLVYKRAEPG-----GPOEPGLASSPFGRA 337
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1176 LTIDVS--AISNLIRKHVSEARLVEDIGHETLVLPYEAKGEGAFVLEHIDRLSDLG 1233
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 338 PLSSCELSQYQFIRKHVASCGLVSDTSTELSTILPSEAKKAGAFRLRQHLERSIDALH 397
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1234 ISSYGISETTELEIFLVAEE-----SGYDA-ETSDGTLP----- 1267
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 398 LSSFGIMDTLEVEFLKYESEDOLENSSEADYKSRDVLPGABGAPSGGAGHAGNARCS 457
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1268 -----ARRRRA-FGDKOSCLRPTEDDAADPNDSIDIPESRETDLS 1309
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 458 ELTOSQASLOSASSVSGARKGDEGAGTYDYGDRPLF-DNQPDP--NSLSLOVEAEALS 514
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1310 GMDGSGYQVKGWMLTQOOYVALLMKRLILARRSKGFEFAOYLPVAVYCAIVAFSLIVP 1369
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 515 RV-GGGSRRKLDGWLKVRQFHLILKRFKRCARNSKALFSLILPAFFVCAMVVALSVR 573
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1370 PRGKTPSELOPWNMYNEQT-----FVSNDAPE-----DTGTELLNALTKDPGF 1414
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 574 ETIDRPLVLSPOYH-NITQPRGNFIRYANDEREYLRSLSPDASPOGLVSTRLSGV 632
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1415 GTRCM-----EGNPT----- 1424
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 633 GATCVLAKSPANGSLGPTLNLSSGESRLAARFDSMCLESFQOGLPLSNFYVPPSPAPS 692
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1425 -----PD-----TROAGEEETATP-VROTIMDLFONGNTMOPNSPACQSSDK 1469
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 693 DSPASPEDQLAANNVSLRPTAGREMTTSPSLRVLREVVR-----CTCSAQG 740
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1470 IKKMLRVCPRGAGLRPPRKONTADILQDLTGRLTSDYLVKTYOQITAKSLKKNIWNE 1529
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 741 TGGS--CPSSVGG-HPQMRVUTGDIIMDTGHNSEVLEFTSRF-----RL 785
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1530 FRIGGSELSVSNQALPRPSEVNDATIKOMKKHKLAKDSSADBEFLNSLGRFTGDLTRNN 1589
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 786 HRYGATFG--NVLKSTIPASFGTRAPRWYK-----IAVRRA 820
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1590 VKVFNKNGHAISSFLNINNALRANLQGE-NPSHYGTAENHPLNLKQOLSEVAL 1648
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 821 AQVYNNKKGHSMPTYLNSLNALIRANLPRKSGNPAAGITVTPNPMKTSASIS-LDY 879
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1649 MTSVDVLVSTCVTFAMSPFASVYVFLIOBRYSKAKHLOFISGVAPVYIWLNSFVWMC 1708
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 880 LLOGDDVIAIFIIVAMSPVSPASVYVFLVAKSTKAKHLOFVGSCHPIIYMLANLYWML 939
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1709 NVVPATVITITICFOOKSYVSSTNLVYALLLXGWSITPLMPASVVFETISTAY 1768
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 940 NLYVATCCVITILFVFDLPATYSTNPAVYLSLFLYGWSITPLMPASVVFETISSAY 999
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

QY 1769 VIFSNIPTGINGSVATVLELFT-DKLNINILDKSVFLIFPHFGLGKIDWVKNQA 1827
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1000 FLIVNLNLTGATATVATLLOLFELHDKDLKVNSTLSCLLIPFNLYLHGLMEMAYNEV 1059
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1828 MADLIERGE-NREYSPLSMDLVGRNLFAMAVEGVFPLITVLIOYFFIRPPRYNAKLS 1886
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1060 INEYIAKIQGPKKMSPEEMDIIVTGLVAMAVEGVGLITIMQVNFLLRPPORMVSTK 1119
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1887 PLNDEDEDVRRRQRIIDOGGONDLLEIKELTKIYRRK---RKPAVDRIQVTPGECFG 1943
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1120 PVED-DVDAVASERORVLRDADONDMVKIENLTKYKSKIRIILAVDRLCGVRGSCFG 1178
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1944 ILGVNAGKSPFKMLTGTPTVTRGDAPLKNKSLISNTHVQNMKGVCPOFDATIELTG 2003
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1179 LIGVNGAGKSTFTFKMLTGDESTTGGEAFVNGHSVYLKELLQOOSLGYCPOCDALDELTA 1238
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 2004 REHEVEFALLRGVPEREKGVEGMARIKLGLVYGEKAGANSGGNKRLSTAMALIGP 2063
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1239 REHLQYTLRLGISMDEARVYKMALEKLETKRYADKPAGTYSGGNKRLSTALALIOY 1298
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 2064 PVVFLDEPTGMPKARFPLMNCALSVKREGRSVYLTSHMECEBALCTRMALVNGRFR 2123
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1299 AFIFLDEPTGMPKARFPLMNCALSVKREGRSVYLTSHMECEBALCTRMALVNGRFR 1358
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 2124 CLGSOHLKRNFGDGYTIVRIAGSNPDLKPVODFFGLAFPPSGVYKKEHRNMLQYOLPSS 2183
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1359 CLGSIQHLKRNFGDGYTIVRIAGSNPDLKPVODFFGLAFPPSGVYKKEHRNMLQYOLPSS 1417
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 2184 LSLARIFSLISOSKRLHIEDYSVOTTLDQYFVAFADQSDP 2227
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1418 HISLAOVFSKMEQVSGVLGIEDISVSTTIDNVPFAKKQSDN 1461
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

RESULT 3

```

ATP binding cassette transporter ABC2 - mouse (fragment)
B54774
C:Species: Mus musculus (house mouse)
C:Date: 23-Mar-1995 #sequence_revision 05-Apr-1995 #text_change 02-Feb-2001
C:Accession: B54774
R:Luciani, M.F.; Denizot, F.; Savary, S.; Mattei, M.G.; Chmelnik, G.
Genomics 21, 150-159, 1994
A:Title: Cloning of two novel ABC transporters mapping on human chromosome 9.
A:Reference number: A54774; MUID:94375008; PMID:8088782
A:Molecule type: mRNA
A:Accession: B54774
A:Residues: 1-1472 <LUC>
A:Cross-references: GB:X75927; NID:q495258; PIDN:CA53531.1; PID:q495259
C:Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homolog
C:Keywords: ATP; nucleotide binding; P-loop
F:44-234/Domain: ATP-binding cassette homology <ABC1>
F:61-68/Region: nucleotide-binding motif A (P-loop)
F:108-1300/Domain: ATP-binding cassette homology <ABC2>
F:1126-1133/Region: nucleotide-binding motif A (P-loop)

```

Query Match 26.5%; Score 3129.5; DB 2; Length 1472;
 Best Local Similarity 46.0%; Pred. No. 4.1e-194;
 Matches 676; Conservative 207; Mismatches 377; Indels 211; Gaps 28;

```

QY 888 MEEPRHKLKGVSIQNLVYVRDGMKVAVDGALNLFEGQITSFLGHNAGKTTMSILT 947
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 16 MEEPRHKLKGVSIQNLVYVRDGMKVAVDGALNLFEGQITSFLGHNAGKTTMSILT 75
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 948 GLPPTSGTAYILGKDIRSEMTIRONLGVCPQHNVLFPMLVEEHIMFYARLGLSKH 1007
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 76 GLPPTSGSATIYGHDIRTMDLRKLGMCPOHNVLFPRLTYVEEHLIMFYARLGLSKH 135
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1008 VKAEMOALDVGLPSSKLSKTSQSLSGGMORLSTVALAFVGSKVYI LDEPTAGVDYS 1067
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 136 IRKETDKMIEDLEL-SNKRHSLVOTLSGKKRKLSTVALAFVGSRAIILDEPTAGVDYA 194
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1068 RRGIMWELLKRYRQRTIILSTHMDADVLGDRITAIISHGKLCVSSILFLKNOLGTGY 1127
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

Db 195 RRAIMDLILKYKGRFTIILSTHMEADLLDRIALIIISHGKLCGCGSPLELKGAYXDGR 254
Qy 1128 LTLVKKDVSSISCRSSSTVSVYLKKEDSVSOSSDAGLGDHSDTLTIDVASINL 1187
Db 255 LTLVQPAPEPTSOEPGLASSPSCGPRLLSCSFPQ-----VSQF 294
Qy 1188 RKNVSEALVEDIGHELTYLVPEAKKGAFFVELFHEIDRLDLSGISSETLEET 1247
Db 295 RKNVASSLVSDETSELSTIIPSAVKKGAERFQOLESIDLALHLSFGIMDTLEEV 354
Qy 1248 FLKAAEF-----SGVNA-ETSDGTLPARNRNRAFGDKOSCLPFTF----- 1287
Db 355 FLKVSSEDDSLSENSEADYKESRKDVLPAGAGLTAVGGQAGNLARCELAQSOASLOSASS 414
Qy 1288 -----DDAADPNDSDIDPESRBTDLISGMDKSGSYOKGMKL 1324
Db 415 VGSARGEETGYSDGYDYRPLFDNIQDDP--NVSLOEAEEMELAQV--GOGSRKLEGWML 471
Qy 1325 TQOQFVALLMKRLLIARRSKGFFAQIYLPVAVFCIALFSLVPPGKYPSLEQPMX 1384
Db 472 KMRQFHLVKKRHCARRSKALCSOILPAPFVCAVMVALSVPEIGDLPPLVLSPOY 531
Qy 1385 NEOYT-----FVSNDAPE-----DTGLELINALTKDPGCTRCM----- 1419
Db 532 H-NYTOPRGNFITYANEBQOEFRLLSPDASPOQLVSTFRLPSPGVGATCVLKSPANGSLG 590
Qy 1420 -----EGNPI-----PD----- 1426
Db 591 PMINISGESRLLAARFFDSMCLESFTQGLPLSNFVPPPPSPASDSVYXDEDSLOAWN 650
Qy 1427 --TPCAGEEWTAP--VPQIMDLFONGNMTMOMPACOSSSDKIKKMLVCPGPGAG 1483
Db 651 MSLPPTAGETWTSAPSLPRLVHEPV-----CTCSAOGTGFS--CPSSVVG 695
Qy 1484 LPPPOKONTADILODLTGRNISDYLVYVOIIAKSLKNIWNEFRYGEFSGVNTQ 1543
Db 696 -HPQARVYTGILTDIGHNVEYLFTSDRF-----RLRRYGAITFG--NVQ 741
Qy 1544 ALPPOEVDATIKOKKHLKAKSSADRLNSIGREPTGIDTRNNVYKWFNKKGMHATS 1603
Db 742 KSIPTAS-----FEARPPVWRKIAVRAQVLYNNKGYHSM 778
Qy 1604 SFLVNNALIRANLOKE-NPSHYITAEFNHPLNLTQOOLEVALMTTSVDLVASICY 1662
Db 779 TYLSNNALLIRANLPKSKGNDAVYXITVNNHPRKNTSASLS-LDVLLOCTDVVIAIFII 837
Qy 1663 FAMSFPASFPVFLIOERVSXAKHLOFISGVKPYIWLNSFWMDKCNVYPATVLIIFI 1722
Db 838 VAMSEFPASFPVFLVAKSKAKHLOFISGVKPYIWLNSFWMDKCNVYPATVLIIFI 1722
Qy 1723 CEQOKSVSSINLPVALLLILYGSITPLMPASFVFIIPSTAYVVLTSVNLFTIGINS 1782
Db 898 VFDLPATFTSPNFPAYLSLFLYGSITPLMPASFVFIIPSTAYVVLTSVNLFTIGINS 1782
Qy 1783 VAFVLELFT-DNKLNNINDILKSVFLPHHCGLGGLIDMKKNQAMADALERFGE-NRF 1840
Db 958 VAFVLELFT-DNKLNNINDILKSVFLPHHCGLGGLIDMKKNQAMADALERFGE-NRF 1840
Qy 1841 VSPSLMDLVCRNFAMAVEGCVFLLTVLLOYRFTIRPPVAKISPLINDEDEVDARRRQ 1900
Db 1018 KSPFEMDITVIRGLVAMTVGFGFLLTMCQYNFLRQGRPLPVSKRPED-DVDVASRQ 1906
Qy 1901 RIIDGGGONDILEIKELTYIRRK--RKPANDRICVGI-PPGECFGLLVGAGKSSST 1956
Db 1077 RYLRCGADNDMKIENLTVYVSKRIGRLIADRLCLGVCPGRCFGLLVGAGKSSST 1956
Qy 1957 KMLTGTITVTDADPLNKNLSLNIHEVHONGYCPQPDATTELTLTGEHNEFFALLRGV 2016
Db 1137 KMLTGTITVTDADPLNKNLSLNIHEVHONGYCPQPDATTELTLTGEHNEFFALLRGV 2016
Qy 2017 PEKEVGEVEMIRKLGIVYGEKAGYSGGNKRKSLTAMALIGGPVYVLEDEPTGMD 2076
Db 1197 PKKDAOVVAKALELTLTADKADKAGYSGGNKRKSLTAMALIGGPVYVLEDEPTGMD 1256

```

```

Qy 2077 PKARFLMNCALSVYKGRSVLTSHSMECEALCTRMALNVGPRCGLSGVOHLKNRFG 2136
Db 1257 PKARFLMNLTLIDIKRGSRVLSHSMECEALCTRLALNVGGRJHJGSLIOHLKNRFG 1316
Qy 2137 DGYITVIRIAGSNDLKPVODFGLAFPGSVLKEKRRNMLQYOLPSSLSLARIFSIJSO 2196
Db 1317 DGYITVIRIAGSNDLKPVODFGLAFPGSVLKEKRRNMLQYOLPSSLSLARIFSIJSO 2196
Qy 2197 SKRRLHEDYSVSGTTLDOYFVNFARQDSND 2227
Db 1376 VGVLTGIEDYSVSGTTLIDNFVNFARQDSND 1406

RESULT 4
Probable ATP-binding cassette transporter ABC-3 - human
N:Alternate names: ATP-binding cassette transporter ABC-C
C:Species: Homo sapiens (man)
C:Date: 29-Jan-1998 #sequence_revision 06-Feb-1998 #text_change 02-Feb-2001
C:Accession: S71363
R:Klugbauer, N.; Hofmann, F.
FEBS Lett. 391, 61-65, 1996
A:Title: Primary structure of a novel ABC transporter with a chromosomal localization
A:Reference number: S71363; MUID:96326608; PMID:8706931
A:Accession: S71363
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-1704 <KUD>
A:Cross-references: EMBL:X97187; NID:91514529; PIDN:CA65825.1; PID:e243436; PID:9151
A:Experimental source: cell line medullary thyroid carcinoma
A:Genetics:
A:Gene: GDB:ABC3
A:Cross-references: GDB:3770735; OMIM:601615
A:Map position: 16p13.3-16p13.3
C:Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homolog
F:255-283/Domain: transmembrane #status predicted <TM1>
F:307-329/Domain: transmembrane #status predicted <TM2>
F:345-364/Domain: transmembrane #status predicted <TM3>
F:373-394/Domain: transmembrane #status predicted <TM4>
F:401-422/Domain: transmembrane #status predicted <TM5>
F:452-475/Domain: transmembrane #status predicted <TM6>
F:549-739/Domain: ATP-binding cassette #status predicted <ABC1>
F:565-690/Region: nucleotide-binding motif A (P-loop)
F:685-690/Region: nucleotide-binding motif B
F:1100-1120/Domain: transmembrane #status predicted <TM7>
F:1145-1169/Domain: transmembrane #status predicted <TM8>
F:1181-1207/Domain: transmembrane #status predicted <TM9>
F:1245-1264/Domain: transmembrane #status predicted <TM10>
F:1299-1324/Domain: transmembrane #status predicted <TM11>
F:1399-1590/Domain: transmembrane #status predicted <TM12>
F:1416-1423/Region: nucleotide-binding motif A (P-loop)
F:1535-1540/Region: nucleotide-binding motif B
F:674,866,1524/Binding site: phosphate (Ser) (covalent) (by cAMP-dependent kinase) #status pr
F:1344/Binding site: phosphate (Thr) (covalent) (by cAMP-dependent kinase) #status pr

Query Match 22.4%; Score 2645.5; DB 2; length 1704;
Best local similarity 35.7%; Pred. No. 1,4e-162;
Matches 639; Conservative 300; Mismatches 523; Indels 327; Gaps 48;

Qy 534 AGIVETGTPGSI-E-LPHNVYKIRMDIDNERTN-----KIKGYW----- 574
Db 131 AAUVFHEFNSKEPLPLAVKYHLRF--STRNNYMTOTGSPFLKETGTGHTTSLFPL 187
Qy 575 --DPPRADPDED--MKYVWGFAIYQDVVEQATIVVLGTG-----KKTGYVQOMPY 624
Db 188 PNPSPREPRTSPDGEPEYIREGFLAVQHAVDRAIMYHADATROLQFLVYTIKRPY 247
Qy 625 PCYVDLFLRVMSRSMPLFMTLAWISVAVIINGIYERKARKETMRINGDNLSTMP 684
Db 248 PPIADPFLVAIQYOLPLLLSTFTYALITARAAYOEKRRILKEYRMWGLSSWLHWSA 307

```



```

OY 685 WFISSLPLVNSAGLLVYLKLG-----NLPSDPVSVFVLSVFAVYTLQCELI STL 739
Db 308 WFLFPLFLVLIASFMTLFLCVKPKPNVAISRDPISLVFLFLOCPALSTISSEFMSTF 367
OY 740 FSRANLAACGGIIFETTYLPVLCVAMODVVGFTIKTFASLSPVAFGCEYFALFEE 799
Db 368 FSKANMAAFGGELFFETTYIPFPAPRYNNMTLSOKLSCCLSLSNVAMAMAOOLIGKEEA 427
OY 800 OGIGVOMONLFFSPVE-EDGENLTTSISMFLFTELYGVMTWYEAFFPGOYGIIPRMFY 858
Db 428 KMGVQIOMDNL-SPVNVVDDCEFGVIGLMLLDSVLGLVWYENAVFPQGFPGOWPT 486
OY 859 PCKTSYWFGE-----ESDEKSHPGSNOKRMSLCEMEEPETHLKLVSITONLYKVRIDGM 912
Db 487 FIMPSTWCGKPRVAVAGKEEDSDP---EKALRNYFAPEBEDVAGIKIKHLSKVRVGN 543
OY 913 K--VAVDGLANFTEGQITTSFLGNAGAKTTMSILNGLPPTSGTAYILGKDIRSEMST 970
Db 544 KDRVAVRDNLNLYEGQITVLLGNAGAKTTMSILNGLPPTSGRAYISGYELSDQMOY 603
OY 971 IRONLGVCPQHNVLJEDMLTYEENHFWARLKGISEKIVKAKEMQALDVGPSKLSKTS 1030
Db 604 IRKSLGLOPOHDLFDNLVTAHEHYTAQLKGLSRQKCEYVKQMLHIGL-EDKWNRSR 662
OY 1031 SOLSGMOKRISVALAFVSGSKVYLDEPTAGVDPYSRGIMELLYKRGRTIILSTHH 1090
Db 663 RFLSGMRKRKISGIALIAGSKVYLIDEPISGMDAISRAINDLOROKSDRTIYLTHTF 722
OY 1091 MDEADVIGDRIATISHKGLCCVSSIFIKNOLGTYLVLVKDVESSLSCRNSSSTVS 1150
Db 723 MDEADLIGDRIATIMAKGELCCSSIFLKKOKYAGYHMLVKE-----P 766
OY 1151 YKKEDSVSOSSDAGLGSDESIDLTDVSAISNLRKRVSEARLVEDIGHETLYVLPY 1210
Db 767 KQNEP-----ISQLVHHVHNALLESAGELSTFLIPR 800
OY 1211 EAAKEGAFVELFHEIDRLSDLGSSYSITTEIFLKAFA--ESSVDAETSDGTLPA 1268
Db 801 ESTHR--FESLFAKLEKKKEKIGIASFGASTTMEVFLRWKGLVDSMDIQAIQ--LPA 856
OY 1269 ---RRRRKAFG---DKOSCLRPFTEDDADPNP---SDIDESPRTDLSDMDCKSGSYQV 1319
Db 857 LOYOHERRRRSDMAVDSNLC-----GAMPDSDIGALIEERRKAVKNTL----- 901
OY 1320 KGMKLTQOQFVALMKRLIARSRKGFACQVLPVAVVCAIYVSLVPPGKYPSLIEL 1379
Db 902 ---ALHCOQFMAMFLKKAAYISWRKKMAYAAOVLYPLTCTVLAL----- 942
OY 1380 QFWMINEQYTFEVSNDAPEDTGLLELNAITFYDPGFGTRCMEGNPIPDTPQAGHEEMTVA 1439
Db 943 -----AINVSELPDDPML--RLTLG-----EYGRP 966
OY 1440 PVYPTIMDLFONGMNTMNPSPACQSSDKKMLPVCPCGAGLPPRQKONTADILQD 1499
Db 967 VPFESVPGTSQIGQOLSHLKDALQAEQ-----QEPREVLGD 1003
OY 1500 LTGRNISDYLVKYVQVLTAKSLKNKIWNERY---GGFSLGVSNTQALPSPQOEYNDAL 1555
Db 1004 L-----EEFLI-----FRASVEGGGN----- 1020
OY 1556 KOMKHLKLANDSSADRFNLNGLGFMGTGLDTRNNVYKFMNKGWMAISSFLYANNAILR 1615
Db 1021 -----ERCL--VAASFVDGERTVUNALFNGQAYHSAPATAIYAVDMLLTK 1063
OY 1616 ANLQGENPSSHGTTAFNHP-----LNLTKOOLSEVALMTTSVDLVLSICVIFANSEFVA 1670
Db 1064 --LTCG---PHASIVSNFPQPSALQAAKQDNE---GRKGFDALNL--LFMAAFIAS 1113
OY 1671 SEVVFLLQERYSKAKHLQFISGVKPVYIYWLNSFVMDCMYVNPATVITVITFCFOOKSYV 1730
Db 1114 TFSILAVSERAYOAKHVQSVGVHVASFWMLSALMDLISFLIPSLILLVVFRAFDYKRAFT 1173

```

```

OY 1731 SSTNLPUVALLILXGMSITPLAMPASFVKIPSTAYAVVLTSLVNLFIGINGSVATFVEL 1790
Db 1174 RGDHNAADLLILLILGMAIIPPLMYLMEFFLAGATATRLIRINLSGT-----ATFLMYT 1229
OY 1791 ---FTDNKLNININDILSKSVLIPPHFLGRLIDMKVN-----QAMADALERFG- 1836
Db 1230 IMRITAVALEELSKTLDFVFLVLPNHCGLMAVSSFEYENYETRRYCTSSSEVAAYHCKYNI 1289
OY 1837 ---ENREVSPLSMDL--VGRNLFMAVEGVFFLITVLYIQRFIFPRPNKAL----- 1885
Db 1290 QYQENFY-----AWSAPGVGRFVSAASGAYILILFLIETNLLQRLICLARRRRL 1345
OY 1886 -----SPLNEDEDVREORORILIDGGQNDT---LEIKELTYIKRRKP--AVDRICV 1934
Db 1346 TELTRMPVLPEDDVADERTRILADSPSLHTPLIKELISVY--EQRPVLLAVDRSL 1404
OY 1935 GIPPEGCEGLLVNGACKSSTFMKLTGDTTVTRGDAFLKKNSTLSIHVHQMGCYQF 1994
Db 1405 AVOKGECFGLGFGAGKTTTFKMLTGESLISGDAFVGHNISSDVGVKVRORIGYCPQF 1464
OY 1995 DATFELLTGRHEHFAFALLRGVEKEVKGVEWALRKLGLVYKERYAGN SGNKRRKLS 2054
Db 1465 DALIDHMTGHEMLVYARLRGLIPERRHIGACVENTRLGLLEPHANKLVRTYSGNKRKLS 1524
OY 2055 TAMALIGPPEVFLDEPTTGMDPKARFLMNCALSVYKGRSVYLTSHSMECEALCTRM 2114
Db 1525 TGIALIGEPVAVIFLDEPSTGMDPVARRLMDTVARARSGKAITTSHSMECEALCTRL 1584
OY 2115 AIVWNRFRCLGVSQVHLKRRFGDGYIVRT--AGSNPDLKRVDDFFGLAPGSVLEKH 2172
Db 1585 AIVWQGFQCLGSPHUKKSKFGSSYSLRAKYQSEQOJALEFFAFAVDLPFGPSVLDEH 1644
OY 2173 RNNLYQVLPSSLSLARISLSQSKKRLHEDYSVQSTITDOYFVNPA 2221
Db 1645 QGVVHYHLRGRDLSMAKVGILKAKKRYGVDDISVQISLEQVFLSPA 1693

RESULT 6
T33783
hypothetical protein Y398C.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 17-Mar-2000
C:Accession: T33783
R:Becker, M.; Graves, T.; Yoakum, M.
submitted to the EMBL Data Library, October 1998
A:Reference number: Z21408
A:Accession: T33783
A:Status: Preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-1802 <BEC>
A:Cross-references: EMBL:AF101313; PDB:1AC69223.1; GSPDB:GN00023; CESP:Y398C.1
A:Experimental source: strain Bristol N2; clone Y398C
C:Genetics:
A:Gene: CESP:Y398C.1
A:Map position: 5
A:Introns: 45/3; 114/1; 195/1; 230/3; 543/3; 794/1; 849/1; 1036/2; 1099/1; 1132/3; 11
C:Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homolog
Query Match 17.5%; Score 2060; DB 2; Length 1802;
Best Local Similarity 27.9%; Pred. No. 1,6e-124;
Matches 551; Conservative 388; Mismatches 630; Indels 454; Gaps 52;

OY 385 YTPDPATRGVMAEVNKTPOELAVFHDLGMBEILSPKWTMTMENSQEMDLVRLMDSRD 444
Db 119 YAPTDTATKQIMKIQNRY-----TANDL-----LNP-VAALIKGLVNTAVPVLNTNMT 167
OY 445 NDHFWEQDLGDLDWTADIVAFIAKHPEDVOSSNGSVYTRAFETNQAIRTISRFECC 504
Db 168 YKGF-----TTBEHNSWMOGQOS-----EC 189
OY 505 VNLNKLPIATEVWLINKSMELDERKFPAGIVFTGIPGSLERPHVKKYKIRMDINVE 564

```

Db 190 DN-----PLL-----AGIVDDSIADKLNPKDRFTYTRLNTH 225
 Qy 565 RNNKIKDGDWDPGRADPEMDKRYWG-----GFAY 595
 Db 226 RNR-----NAFGDSYPMDSVSVSAVGYSGPIINPDNDGSGCYWQGEFMT 273
 Qy 596 LODVBOAIRVLTGTEKK-----TGVMQOMPYPYVDIIFLRVMSMPLFMTLAMI 650
 Db 274 VQAVAVALTTEITGEDAQLTELLDSYQSRFPFGYSTKI-IEIGAFMVAIVIFESMT 332
 Qy 651 SVAVITKGIYEEKARLKTETMRIMGDINSILMFMSISLPLVASGLVLIKLGNL 710
 Db 333 SVIYIRAVAVKEEDLKEKYMKGISOFTINVAHFIIINAKLTFRAVIVLIIMHFAL- 391
 Qy 711 PYSDPSVFEVFLSVFAVVTILQCFILSTLSRANLAACGIIYFTYL----- 759
 Db 392 -KSDMLTFMFLMIAFDVYVFAFMISFMNSATSATLISVFMMLLFVWAFSSIDOT 450
 Qy 760 -PYVLCAVADYVGFLLKTFASLSP-VAFGCEYFALFEEOGIGYOMDNLFESEVED 817
 Db 451 NPYPL-----GYRL---INCINDIALNGLQLLAAYEQADGLKMGELFTPSFDN 499
 Qy 818 GFNLTSISMPLFDTFLYGVMTYEAFFPGQYGP-RPMYFPCYSYWF---GESEDEK 873
 Db 500 NLFFGHALLIYDGIIMILTYIEAVIRGGSGVPOKPMF-VLPSTWFPNPGSKTYDS 558
 Qy 874 SHPGSNOKRMSSEICMEDEPTHLKGVSIONLVKVY-----RQMKYAVDCL 919
 Db 559 SDQOQOIQYADHYKLEKEPTDLPTINVNLTKTYGTFEKKLFEDCKFGSGKRAVSNL 618
 Qy 920 ALNFEQGISFLGNAGKTITMSLTGTFPTSGATYLGKDIRKSEMTIRONGVCP 979
 Db 619 NLKWPPOCVLLGNHAGKSTFTSMLTGVAOSSGSAVYNDITSLPKIRREGICLP 678
 Qy 980 QHNVLFDMLTVEEHITWEYARLKLGS-EKHVKAMEQOMALDVLPSSKLKSTQLSGM 1037
 Db 679 QYNLFGMTYMEHLFEPAKLEKERTWDPPEARLRLRIDF-----KAFPMAGALSQGO 733
 Qy 1038 QRLKSVALLAFVGSKVYILDEPTAGVDPYSKRGIMWELLKYYROGRTIISTHMDADV 1097
 Db 734 KRKLSTLALILGSEVWMLDEPTSGMDPGARHETWTLQKEKERTILLTHHMERDIL 793
 Qy 1098 GBRILIIHSGKLCGSSSLFLKNOLGTYLLTLVKKDVESLSSCRNSSSTVSYLKKEDS 1157
 Db 794 GBRILIAHGOLECCGSMFLKQGGYHLFIY----- 827
 Qy 1158 VSQSSSDAGLGSDBSDTLTIDVSAISNLKHYSEARLYVDIGHETLYLPEAKEGA 1217
 Db 828 -----YDTTSTPDVSKTDTIIREYIPEAHVESYIQEATYLL-SATHRPI 871
 Qy 1218 FVELHEIDRLSDGISYSIGSETLEIFLKVAE-----ESGVDAETSDGLPAR 1269
 Db 872 FPLFELEEDHOTCGITISFGVSTTMEVFLKVGHTADERYNEHGIENDISE----- 925
 Qy 1270 RNRRAFGDKQSLRPTEDDAADPNDSOTDPESRETDLLSGMGKSGYQVKGMLTQOQF 1329
 Db 926 -----MIEKD-DPILODL-----RAQRYVGFLLQMOHA 953
 Qy 1330 VALMKRLLIARSRKGFPAQVILPAVFCIALVSLIVPPCKYPSLELOPMWYNEQYT 1389
 Db 954 KAFYFRAIFEFKKTQFPLQVFPVAYIVLAVFTSOVLPSVKE----- 997
 Qy 1390 FVSNDAPEDTGLLELLNALTKDPGEGTRCMGNPIPDTPCQAGEEWTATAPOTIMDLF 1449
 Db 998 -----QDPQITSLA-----PFSDFK-KAG-----HLVSD- 1020
 Qy 1450 QNGNMTPONPSPCQCCSDDIKKMLPVCPPGAGGLPPQPKRQNTADIIQDLTGRNISDY 1509
 Db 1021 -SGNYVTL-----LGG-----SOMLSMV 1038
 Qy 1510 VKTYVOALIAKLNKIMVNEFRYGSGLGVSTQALPSPQOEYNDALIKOMKHLKIAKDS 1569
 Db 1039 QGIYTO-----LGYTO-----VVDITSNVEKFTIMQUTNAM 1069

Qy 1570 ADR-----FLNSLGRMTGLDTRNNVKNVKNNGKMAISSFLVNNAILRANLQK 1620
 Db 1070 GSRFTGLHAYALCFVPSMNFST--VSVPSLKTFEPPNNGGLTALAIPTDSMILSKQK 1127
 Qy 1621 GNPMSHYGTAEHNPBLNTKQOLSEVALMTTSVDVLYSICYIFAMSVVPSVFFLIQER 1680
 Db 1128 -----QSFYAVNHHLPLPSTODTLKNTNRSQAAFLAYGLVSAFVAVGVISOFLTER 1182
 Qy 1681 VSRKHLQFISGVKVIYVWLSFVMDKCNVYVPATVLIIFICFOOKSYSSSTN-LPULA 1739
 Db 1183 KRKSKHOLLGSRPMWELTAFIWDAAFYRILCFPAIFYIININATYHDFCVMLIT 1242
 Qy 1740 LLLILGWSITPLPMPASVFKIPSTAYVLTYSNLTGINGSVATVLELFTONKLNI 1799
 Db 1243 LSLFLKGMALPPTYVFOCFEESAPRGFMVYMHILGMISIAVPIISOTSSIDAGYL 1302
 Qy 1800 NDILKSVELIFPHFCGRGLIDMKKNQAMADLSEF-----GENRFS 1842
 Db 1303 WSIIFA-WLFPYTNISQIATVTFQENENYIACKKLDCTIPMKFNAVTCGCTASRLYVD 1360
 Qy 1843 PLMSDLVGRN-----LFMAVEGVFFLITVLIQYRF-----IRPRYNAKLSPLND 1890
 Db 1361 NVLF--VGNRKGLVYVIFLAVOGFIYMWIMVRENDQTKLALIRCKADNPIMDITD 1418
 Qy 1891 -----EDEDVRRER--QRILDGCGQNDLLEIKELTKIYRKKRPAYDRICVGIIP 1938
 Db 1419 TDKVDERVDSDVIAEKSVYQRL--ANNKTAIVSNLVKVMGNFN--AVKGFVHNS 1474
 Qy 1939 GEGCGLLGVNAGKSSYFKNLTGDTTGTGDAFLKNSILSNHEHYOMGCPQOFDAIT 1998
 Db 1475 KDCFGLLGVNAGACTSTFQOMLTGNSISGDATVYNGMSVKNMRREGANTGYCPOYDAII 1534
 Qy 1999 ELTGREHVEFALLRGVPEKEVKGEMAIRKLGLVKYGEKAYNSGKNRKRLSTAMA 2058
 Db 1535 KEMSGEELIYMFARICRIPKEDIPKKNVAVIHAIGIMYASROIKYSGNRRRLSLGIA 1594
 Qy 2059 LIGGPVYFLDEPTTGMDPKRARRLWMCALSVYKEGSAVLTSHSMECEALCTRAIIV 2118
 Db 1595 IVGLPDVLLDEPTSGVDPKARRIIMILNRDLGLVALVLTSHSMECEALCTRAIIV 1654
 Qy 2119 NGRFRCISVOHLKRNRFQDGYTIVRIAGSNPLDKPVQDFGLAFPGSVLKEKHNMLQY 2178
 Db 1655 YKFCFYCSOCHISRSQSGYTLILRLKNNR-DAEKTFTKQTFGRSGYVKEEHHYQLMF 1713
 Qy 2179 QLPSSLSLARIFFLSQSKRLHIEDYSQSTPLDQVFNPAKMD---GSDDD 2228
 Db 1714 DIPRQDSMSRLEFKLETVSTLSLWMDYSLSQTLQVOVLEFSKDGVSDDSE 1766

RESULT 7
 A:4845
 Probable ABC transporter (imported) - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
 C:Accession: A84845
 R:Lin. X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.
 M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon,
 euss, D.; Niernann, W.C.; White, O.; Elsen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
 Nature 402, 761-768, 1999
 A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
 A:Reference number: A84420; MUID:20083487; PMID:10617197
 A:Accession: A84845
 A:Status: preliminary
 A:Molecule type: DNA
 A:Restrictions: 1-1816 <STO>
 A:Cross-references: GB:AE002093; NID:96598351; PIDN:AAC02761.2; GSPDB:GN00139
 C:Genetics:
 A:Gene: AE2941700
 A:Map position: 2

Query Match 16.8%; Score 1984; DB 2; Length 1816;
 Best Local Similarity 28.8%; Pred. No. 1.4e-119;

Matches 584; Conservative 282; Mismatches 592; Indels 572; Gaps 60;

QY 489 NENNOAIRIS-RFMCVNLKL--EPLATEVWLKSMELLDERKFAIVFTGTPGS 545
 Db 59 DETNNMIDILSLKFPDLRLVTKIFKDDIELETITSANHYGCSEVRNCSNKIGAVVH 118
 QY 546 IELPHVKKIKIM-----DIDNEKRNKIKDGVWDGPRADPE-----DMRY 588
 Db 119 EGGPHLFDSIRLNTMAFAGFPNKSIMDN-----GPIINDLEMOINTIPMOY 169
 QY 589 VNGGAYLADVDEQAIIHVLGTEKKTGVYMOQMPYVDIDILRYSRSMPLPMTLAW 648
 Db 170 SEGGELTLOOVVDSFI-----FASQON-----NDPLSINSLSALREFLPW 212
 QY 649 IYSVAVITKGIYKEAKLKEKTRIMGIDNSILMFMSISLLPLVLSAGLLVVLTKLGN 708
 Db 213 TLFSSSVIRMPFPFRTTDE-----FOSIVKSVMLF-----246
 QY 709 LIPYSDPSVVFELSVFAVVTILOCFILSLFSRANI/AACGGIYFTFLVYLCAVMO 768
 Db 247 LFKYSDKTLVFTFLFGLSALMSLPMISTFTAKTAAVAGTITLPGAFFPY--TVND 304
 QY 769 DYVGFTLIFASLLSPAFAFGCEYFALFEBOGIGVQMDLFSPEVEDGFNLJTSISM 828
 Db 305 ESVSVNLKVYASLSPFAFALGSTINFADYERAHVGLRWSNIMPA-----SSGVSPVCLLM 361
 QY 829 LFDFTLVGMTWYLEAVFPGQYGIKRPWYFPTKSYWFGESSEKSH--PG-----877
 Db 362 LIDSTLYCALGLYLDKVLPRNGVRYPMNFTSK--YFGKKNKNDONRLPGETDMPAD 419
 QY 878 -----SNOKRMSTI--CMEEPTHIKLGVSIONLVKY--RDGKAVAV 916
 Db 420 IENVGEPFDPVESISLEMKOELDRG-----IOVRNHLKYYASRRGCCAV 468
 QY 917 DGLALNFESQITSLFNGAGKTTMSILTGLPPTSGTAATLGDISESTIRONIG 976
 Db 469 NSTDLTYENQOISLICHNGAGKSTIISMLYGLPPTSGDALILGNTITNDELRKEIG 528
 QY 977 VCPQHNLFEMLTVEHILMYAALKLSKHYKAEMKONALDVGFPSSKLKSTQSLSG 1036
 Db 529 VCPQHDLFELTVEHLEFAVLLKGVESSEKSTVVDMAFEVGL--SPKINTLVALSGG 587
 QY 1037 MOKRLVALAFVGSKVLLIDEPYTAGVDPYSRGIMELLKXROGRTIILSTHNDADV 1096
 Db 588 MKRRLSIGLIGNSKVILLDEPTSGMDPYSMRLTMOULKTKRITILLTHSHDAEE 647
 QY 1097 LGRRIALISGKLCCVGSILFLKNOLGTGYLLTVLKKDVESLSLSCRNSSSTVYLKED 1156
 Db 648 LGRRIAMANGSLKCCGSSIFLKHNGVGYTLTVK-----TSPT-----687
 QY 1157 SVSSSSDAGLSDHSDPTLTIDVSAISMLIRKHSEARLVEDIGHETLYVLYPEAKEG 1216
 Db 688 -----VSAVAIYHRHPSATVCSEVGENESFELP--LASLP 722
 QY 1217 AVEVLFHEIDRL---SDL-GISSYGISETTLIEFLKVAESGVDAETDGLPARRN 1271
 Db 723 CEENMRREIESCMKNSDSDPGICQGISVTTLEVLARA--GCNLDL-----770
 QY 1272 RAFFGKQSCLRPTEDDAADPNDSIDPESREDLLSGMDKGSYVK-----GW 1322
 Db 771 -----DKO-----EDLEFVSPDKSSLYVCGSNOKSMOPKLIASCNDAAGV 811
 QY 1323 KLTQ-----QGFVALLM-----KRLILARRSKGP 1347
 Db 812 IITSYAKAFRLIAVAVMTLLIGFISIOCGCSIIISRMFWKRAKALFKRRBSACRDKTY 871
 QY 1348 FAQIYLPVAFVCIALVPSLYPPFGKTPSLFLOPMATNEQYTVSNDAPEDTGLLELNA 1407
 Db 872 AFQITIPAVFLFGLF-----LQTKP-----HPDQKSTITLTAVFNP 909
 QY 1408 LTKDGFSTRCEGNPIPTPCQAGEEMWTAPVQPIINDLFONG-----NMTMGNPSP 1461
 Db 910 LLSGKG-----GGGPIIPD-----LSVPIAKVAVOYIEGMIQPLRNTSYKFPNP 954

QY 1462 ACOCCSDKIKMLPVCPPGAGLPPOROKNTAIDLODITGRNISDYLVKTYVOIIAKSL 1521
 Db 955 -----KEALDAI--DAAGPIGLPILLS-----975
 QY 1522 KNIKWNEFRYGGFSLGSYNTQALPPSGEVNDALQKKKKHLKIAKSSADRLNSIGREM 1581
 Db 976 -----MSFEIWSFD-----QSYOSSRE-----GLSSHSCNHPDGSIG---1009
 QY 1582 TGLDRNNVKKWNNKGMHAISSEFLANYINNALIRANLOKGENSHYGIAPNPLUTK- 1640
 Db 1010 -----YVILHNGTOCHAGPIYINVMHALLR--LATON--KNMTIOTRNHPLPPTKT 1057
 QY 1641 QOLSEVALMTTSVVLVLSICVIFAMSVVPVAFVFLIOERVSAKHLQFISGVKPYIWL 1700
 Db 1058 QIORHHDADFASAIIVNI-----AFSPFASFAVPVIVKEVEYKAKHQQLISGVSLSVYL 1113
 QY 1701 SNFVWDMKNVVPATLVIIIFICFOOKSYVSTNLPVLLILLILYGNSTIPLTPASVFE 1760
 Db 1114 STYVWFISFLPSTFALILFYAFGLEQFIGIGREFLYMLLELGLAIASSTYCLTFEF 1173
 QY 1761 KIPSTA-----YVVLVSNIPT-----GINGSVAFFVLELF--TDNKLNN 1798
 Db 1174 TEHMAQATSSYSVLPILSLFVSFSSNVLLVHFFSGILLVWISHWGLIPATASMSY 1233
 QY 1799 INDI-----LKSVELFPFPCIGRGLDM--YKNOMADALERPGENRVSLSMDLGR 1851
 Db 1234 LKELIFRYALQNFPRLSPPCFCSDDLALIRQMKDKSH-----GVFEMWVGA 1286
 QY 1852 NLFMAVEGVFFLITVLYOYREFIRPRVNAKS-----1886
 Db 1287 SICYLGEVLEFCRYSMLLSFF--HGIDTKLSITVIGASRLTELIDRVYSTSFST 1343
 QY 1887 PL-----NDEEDVREORIIDGGONDILEIKELTKYRKR--KAVVD 1930
 Db 1344 EPLIKDSTGISTDMDDLDIDQEBRRTVSGISDMTMYLQRLKRVYRGDHNHGRKAVO 1403
 QY 1931 RIVCGIPRECEGRLGCVNAGKSTFPMKTGPTVTRDARLKNLSLNIHEHONMGY 1990
 Db 1404 SLFVSVAOACGCFGLTNCAGKTTLSMLSCHEPPTSGATFPGKDIVASPKALRQIHGY 1463
 QY 1991 CPOFDAITELTGREHVEFFALLRGVPERKGVGEAMIRKLGVLKYGKAGYSGCNK 2050
 Db 1464 CPOFDALFELYLVKEHLIYARIKGVYDHRIDNVYETKLEVFEDLKSHRSKPSFLSGCNK 1523
 QY 2051 KRISTMALIGSPVVEFLDEPTTGMDPKARRFLNN--CALSVKEGRSVYLTSHSPECE 2108
 Db 1524 KRISVALIAMIGDPIVILDEPSTGMDPVAKRFMDVYISRLSTRSGKAVAILLTHSMNEAO 1583
 QY 2109 ALCTRAIIVNCRFGICLSYQHLKNRGC-----2136
 Db 1584 ALCTRIGIMVGRKLCTGSPQHLKTRYGNNHLELVPEYNGVKPNEYSNVELNFCOIIQ 1643
 QY 2137 -----DGYT-----IVVRIA-----2146
 Db 1644 WLFNVPTOPRSLIGDLEVICIGVOSITPDTASASEISLSPMORAKFLGNGORVSTLV 1703
 QY 2147 -----GSPN-----DLKPVDFGLAFPGSVLKEKR 2173
 Db 1704 PPLPEEDVRPDOLSEQLFRDGGIPLPIFAEWMLTKKFSALDSFIQSSPGATFSGNG 1763
 QY 2174 NMLQYLPSSLS--SLARISLISQSKRLHIEDYVSOTTLDOYVNNR 2221
 Db 1764 LSIKYOQPFEGGGLSLADAFGHLEARNKRLGIAEYISISOSTLETITFNHRA 1813

RESULT 8
 T15200
 hypothetical protein F12B6.1 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 17-Mar-2000
 C:Accession: T15200
 R:Pauley, A.; Maggi, L.

submitted to the EMBL Data Library, May 1997

A:Description: The sequence of C. elegans cosmid F12B6.

A:Reference number: 218307

A:Accession: F12B6

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1447 <PAD>

A:Cross-references: EMBL:AF003138; NID:g2088708; PID:g2088709; PIDN:AA54153.1; GSPDB:GN

A:Experimental source: strain Bristol N2; clone F12B6

C:Genetics:

A:Gene: CESP:F12B6.1

A:Map position: 1

A:Insertions: 79/2; 114/3; 177/1; 224/3; 331/1; 345/3; 373/2; 417/2; 464/1; 536/1; 659/2; 6

C:Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homology

Query Match 15.8%; Score 1863.5; DB: 2; Length 1447;
Best local similarity 28.3%; Pred. No. 6.1e-112;
Matches 505; Conservative 283; Mismatches 498; Indels 499; Gaps 51;

QY 571 DGYWDGPRADPEDRMRYVWGFAVLQ-----DVEQAIIIRVLTG-TEK 613
DB 2 DSKSNPSPKRLDLKTYIFGFSFLQGSWPSLEQSKSQKLESDRAIMSELTNOTDA 61
QY 614 KTYGYMOOMPYPVVDIELRVMSRMPLEMTLAWYSAVITIKGYVEKEKRIKPTMRI 673
DB 62 NLGYAAQDEPYPCTVKDTF--NVALEMPRLFLSLIFPSALVKNIYVEKEKIKQOMRA 119
QY 674 MGLDNLIMFWPISLPLVLSAGLVLTIKGNLPLSDPSVVFVLSFAVVTIIQC 733
DB 120 MGLGDAVHFISWGLISLVNLTSLIISITKAKIFDIYDYLFLVLLFLESIAMS 179
QY 734 FLISTLFRANILAAACGGIIFLYLPLVLCVAMQDYGF-TLKIFASLSPAFPGCG- 791
DB 180 IFPSFLFNANLAPATCVLWFVFIFPOLRT--DRISPTENRISLLPPTAMGH-CP 236
QY 792 ---EYFALFEEDGIGVQNDLFESEVEDGENTLTSISMFDFFLGVMTWYIEAVFPG 848
DB 237 KLESEFNAMER---ATMSDLEMMNPVLYGISVLCMIMLVVDVAVFLILAWYISAVAPG 292
QY 849 OYGIKRPWFYPCPKSYW-----FGEESDEKSHPGSQKRMSELCMEEPETHKLG 898
DB 293 DGVGRPLMFPPTLKLYMAAGLYKNVEFEVDDEHEDTIPNSOS-----FSEETNLTL- 344
QY 899 VSIQNLVKKYRQGMKVAVDGLANFEGQITSPFGHNGAGKTTMSILTGLEPPTSGTAY 958
DB 345 -----ADCLNLRLYEQGITGLGHNAGKTTMSILCGIYAPSSGTAK 388
QY 959 ILGKDIRSEMTIRONLGYVCPQHNVLFDMLVTEHIMFYARLGLSEKHYAKEMDQALD 1018
DB 389 IYQDRIETDLRRVROVLGICPOHNVLFSHLIVSEQLRLFAALKGVPDSELTQVDELIAS 448
QY 1019 VGLPSSKLSKTSQSLSGMQKLSVALAFVGGSKYVILDEPTAGVDPFSRRGIVELLKY 1078
DB 449 VSL-TEKANKLASTISGGMKRRLICIGATIGSKRFYIIDEPTAGVDPYARAKQIMWLQRN 507
QY 1079 ROGRTITLSTHMHDEADVLDGRIITAIISHGKLCVGGSLFLKNQOLGTGYLLTVKADYSS 1138
DB 508 KEGTITLSTHMHDEADVLDRIATLS-----ODFE-- 538
QY 1139 LSSCRNSSSTVYTKKEDSVSSQSSSDAGLGDHESDTLLIIVSALSILRKHYSEARLVE 1198
DB 539 ----- 538
QY 1199 DIGHELYVLPYEAKEGAFVELFHEIDRLSDLGISSYGISSETTLEIIFLKVAESGVD 1258
DB 539 ----- 538
QY 1259 AETSDGTLPARRRNRAFQKQSCIRPTEDDADPNDSDIDPESRENDLISGMDGKSYQ 1318
DB 539 -----KPDLL--DGR----- 546
QY 1319 VKGWLQGOQFVALLMKRLILARRSRKGFPAQIVLPAVFCIALVPSLI----- 1367

DB 547 ----RLIQHFALLVCRINYLTKSKRTFLQVILPLFLLAELFVLQVSTARPDLNV 602
QY 1368 -VPPGKYPSLELOPMWYNEQTYVSN-DAPEDGTLELNLALTKDGGFCTPMEG--NP 1423
DB 603 SMPPLPLETSI-----MGNSDFVNSWMTAENSTANDILHAMFSFGCTPRCAKDYPN 657
QY 1424 IDPT-----PCQ--AGEBEWT-----TAPV 1442
DB 658 LDDTMRRLMFNRKRGFGKRNKAPGVDDSDVNEQCGNIGEDDYEDLSMATYNAPIY 717
QY 1443 QTIMDL-----FQNGWMTQNPSPACQSSDKIKKMLPVCPPGACGLPPPRKONTADIL 1497
DB 718 CGCEDGWNCTLEDMKMNEN-----WLRNTTDRI 748
QY 1498 QDITGNISY-LVKYVQILAKSLKNKIMWNEFRYGSLSVSTQALPSPQEVND-- 1553
DB 749 FDLTGRNLQFRLITPFAQLANTA-----PFLGGSLGHVQRA--QSOADIDRSK 799
QY 1554 ----AIQMKKHLKL-----AKDSSADRFLNSLGRMTGLDPRNN 1589
DB 800 RGMLETIKDIAQSMRIILNLTGIEPATPKVLDPPAQNTITLMQVYNDL--LQNLQVREN 856
QY 1590 VKWYFNKKGHHALISFLVNTNNAILRANLQKGNPSHYGITAENHPLNT-KQOLSEVAL 1648
DB 857 VKWYFNKRIWGPPIASNLISNALLKQD-DYALDPEDLGLIMNHNMKNTISQTLQNAL 915
QY 1649 MTSVDVLSICV-IFAMSFVPAFVFLIOEVRYSKAKHLQFISGVKPYIYLSNFWDM 1707
DB 916 KFTQALAVFRITILLVLSMLPAGFTYIYVEDRICHALQOLQGGIRKQYVWVTSYLDIM 975
QY 1708 CNYVPATLVITITTCQOKSYASTNLPLALLLLGKSTPLKYPASFPIKIPSTAY 1767
DB 976 VGGIHP-----RHICNNALHPVLP-CLRLYRRRRNLKLPSSL----- 1013
QY 1768 VLTGSVNLFIGINGSVATFVLELFTDNKLNINDIL-----KSVFLIFPHFCIGRLIDM 1822
DB 1014 -----RARNVDSLCLIPKSLFCGGSJLFCNCMFWLL 1046
QY 1823 VKQAMAD--ALERFGENRFVSP-----LSMDLVGRNLFMAVEGVVFLITV 1868
DB 1047 RRRSLCLDSYHARVAYGSEOMNRDPMINOLPLPSLAPDQGHIMKLFHIVIAITICI 1106
QY 1869 LLOY-RE-FIRPRYNAK-----LSPLNDEDEDYRERORI--LDGGQNDILEKEL 1917
DB 1107 FQMDPEFGVRRRENLDAMMLREPSRCDDEDVYKEROVDAIPMD--SSDNHALIYRNL 1165
QY 1918 TKIYRRKRPAYDRICVGIPECEGGLGVGACGSSPKKLT-----GPTVYRGDAFL 1972
DB 1166 AKAYNPDEL-AVKGISFAVEPECEGGLGLNGAGKTTFFAMLTAKIRPGHISIMQNTRI 1224
QY 1973 NKNLSLISNIEHONNGPOPDATITELTGREHEFPALLRGVPEKEVGKGEWAIRKL 2032
DB 1225 NTGS-FSDVNR-FQOLGCTGPOFDALNMKSTRENKLFYARIGIYPTQIDSIIDRLIAL 1282
QY 2033 GLVYGEKAYAGNYSNGNKKLSTANALJGPPVYVLDDEPTGMDKARFLMNCALSVK 2092
DB 1283 HLRPANTOTSSLSGNNRKLVAVALVQPSLIFLDEPSAGMDGSOQFLMKYIERLCK 1342
QY 2093 EGRSVYLTSHMECEALCRMAIWNGRFRLGSVOHLKNRFGDGYTIVYRIAGSNPL 2152
DB 1343 SGRVAVLTSHMECEALCRMAIMDRGIRICLGGGKLSXYGSGMLTKM-GKDENA 1401
QY 2153 KPYODFFGLAF-POSVLKEKHNMLOYLPSSLSLARFSTLSQ 2196
DB 1402 KEIAGIMRSKLGDSRVEALIHCTIFIHIEQGTASVAVALEYVNO 1446

RESULT 9

C88925 protein F33E11.4 [imported] - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 10-May-2001
C:Accession: C88925

Db	621	GRVTO-----	LGVTOT-----	VVDITSNVEKFI----	MDOT- 647
OY	1571	DRFLUSLRFMFGDTRNNKVMWFNNKGMWAISSFLNVLNNALIRANLOKGNPSHYGIT	1630		
Db	648	-----NAMGSRTFEGI-----	HYALGFV-----	-----PSMFNFS 671	
OY	1631	AFNHPLNLTIKQOLSEVALMTTSVDVLISCVIFPMSEFVSPASFVFLIOERKAKHLOFI	1690		
Db	672	TVSPV-----	SLKISFAVCV-----	AGVSOFLITERKKSKHMOFL 707	
OY	1691	SGVRFVJWMLSNFWMDCNVVVPATLVIIIFLFOOKKSVSSTN-LPVLALLLLLYGKSI	1749		
Db	708	SGIRPWFHMLTAEITWDAWAFVIRILCPDALTIFENTALTATHDEGVMLTLLSPFLYGTA	767		
OY	1750	TEPLWPSAFVFKIPSTAYVVLVSVLNLFIGINGSVAIVLELFTDKNLNNINDIKSVFLI	1809		
Db	768	LPFFWYQFEEFESAPKGEWAVTMYHILTMGSIASVPIISOTSLSDACTYMSIIFA--WL	825		
OY	1810	FPHEICLRGLIDWKKNOAMADALERF-----	GENRVSPLSMDLYGRN 1852		
Db	826	FPYVINSQIATVTFQENENVRACKDKDCTIPEFKAVTACCGTASERYLYVDVJLF--GNR	883		
OY	1853	-----LEAAVGVFELITVLIQYRF-----	IRPRYNARLSPND-----E 1891		
Db	884	KGLIVYVFLAOGFYLWTFMFRENDOFTKLFALIRCKRADNITMDITDVKDERDVE	943		
OY	1892	DEDVRRER---GRILDGGGONDILEIKETETIYRKRKRAPADRICVBIIPGEGCEGLLGVN	1948		
Db	944	DSDVIAEKSVYQRL--ANNKRTALVSNNLVYKMGFN--AKGVNPHVNSKDCFGLLGVN	999		
OY	1949	GACKSTSEKMLGDDTIVYRGDAFLFNKNSILSNIEHVHONMGYCPQPDALTITELTGREHVE	2008		
Db	1000	GAGKTSTFQMLTIGENSISSGDAYVNGMSVKKNNRMEADAGANTGYCPQYDALTIKEMSGRETLX	1059		
OY	2009	FPLALGVEBEKGVGEMATRKGLVYKCGRYAGNVSQGNKRRLSRAMLALGGPPRYVL	2068		
Db	1060	MEARIRGPIPEKIDIPKXNAVIAHAIIGMAYASROQKITSGNKKRRLSIGAIYVGLPBYVLL	1119		
OY	2069	DEPTTGMPDKARFELMNCALSVYKEGSRVYLTSISMEECHALCTRAIMVNGRFRGLGSV	2128		
Db	1120	DEPTSGVDPKARRIIMWNLNLRDLGALVLTSHSDECEALCTETELAIWVYKFCYSGC	1179		
OY	2129	QHLKRRFEDGYTIVIRIAGSNPDLKPVQDFGLGAPGSVYJKEKHNMLOYOLESLSLA	2188		
Db	1180	QHIKRYSQSGYTLIRLKNRN--DAEKTSTIKQTFPGSVIYKEHVLQWLFDPIDRGDWS	1238		
OY	2189	RIFSILISQSKRHLIEYVSOTTLQVFPVNPAD---OSDD 2228			
Db	1239	RLFEKLEIVTSLNWDYISLSQTLLEQVLEEFESRDAAGVSSD 1281			

RESULT 10
F88559
Protein C48B4.4b [Imported] - Caenorhabditis elegans
C:Species:Caenorhabditis elegans
C:Date:10-May-2001 #sequence_revision 10-May-2001 #text_change 17-May-2002
C:Accession: F88559
R:Anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biology
A:Reference number: A750001. WUJID:99069613. PMID:9851916
A:Note: see websites genome.wustl.edu/gsc/C.elegans/ and www.sanger.ac.uk/Projects/C.
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999;
A:Accession: F88559
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1758 <STO>
A:Cross-references: GB:chr_III; P1DN:CAA82384.1; P1D:93875025; GSPDB:GN00021; CESP:C4
C:Genetics:
A:Gene: C48B4.4b
A:Map position: 3
C:Superfamily: Unassigned ATP-binding cassette proteins; ATP-binding cassette homolog

Db 1666 MVVQ-LVQRLDPNSVLTKTSTNKLNLKQIPKEKDCWSAKFEWQALAKDGLKDFIL 1724
 QY 2209 SQTTLDOVAVNFA---KQSDDDHLKDL 2234
 Db 1725 AQSLEETFLRLAGLDEQDLDPSTVEIS 1753
 RESULT 11
 T42749
 ATP-binding cassette transport protein homolog - *Caenorhabditis elegans*
 C:Species: *Caenorhabditis elegans*
 C:Date: 11-Jan-2000 #sequence-revision 11-Jan-2000 #text-change 21-Jul-2000
 C:Accession: T42749
 R:Wu, Y.C.; Horvitz, H.R.
 Cell 93, 951-960, 1998
 A:Title: The *C. elegans* cell corpse engulfment gene *ced-7* encodes a protein similar to A
 A:Reference number: 222259; MUID:98297348; PMID:9635425
 A:Accession: T42749
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-1704 <WUY>
 A:Cross-references: EMBL:AF049142; NID:93172340; PIDN:AAC24116.1; PID:93172341
 C:Genetics:
 A:Note: *ced-7*
 C:Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homology
 Query Match 13.0%; Score 1538.5; DB 2; Length 1704;
 Best Local Similarity 24.9%; Pred. No. 1e-90;
 Matches 484; Conservative 347; Mismatches 675; Indels 441; Gaps 59;
 QY 458 WTADIV-----AFLAKHPEDVOSSNGSY-----TWRAFENET----- 492
 Db 24 WTLPELILPLLGLPLVYLKVNADHTSSPENLYDNFQVKGTVEDVLESNFIKPIYKRW 83
 QY 493 -----QAIRTI-----SRFMECVNLNKL-----EPIDTEVWLINKS 523
 Db 84 CLRSVVVGYTSKDAAKRTVDLKKFAERFQSAKLKLSVNESSEQLLT---VLEND 140
 QY 524 MELIDE-----RKFAGIYFTGITPGSTIELPHNYKKIRMDIDNVERTNKTIDGWDGP 578
 Db 141 LPLNLTFCALNSYAGVGVFDEVDTNKKL---NYRILKGTPEETVHILRETSYNDYGP 196
 QY 579 RADPEEDM---RYWVGFAVLDQVVEQALIFVLGTCKKIGVMOQPRYCYVD----- 629
 Db 197 SSGRYSRIPSPPYWMTSAFLTRQHAIESFLSSVQSGAPDLPITLRGIPERYTSSVSA 256
 QY 630 --DIFLRVSRSMPLPMLTATYISAVVITIKGIYVEKEARLKETMRINGLDSILMFSWFI 687
 Db 257 FLDFE-----PTMAFVETFINVITREIAENHA-VKPYLTAMGLSTPMFYAHV 307
 QY 688 SSLILVLSAGLLVILKLGMLLPSPDSVVFELSAFVATVILQCLISTLSRANLAA 747
 Db 308 MAFLKFVYI--FLGSIILPTVMEFVSPALITVLMIGAVLFGAFVAFVANNNSAL 365
 QY 748 ACGGIITFLIPVLCVAMODVYGLTKIFASL--LSPV-----AFGCGEVEAL 796
 Db 366 K-----ALIVAMGAMIGISYKIRPELDQISGFIYGLNGLNMFALAVANISD 412
 QY 797 FEEDGIGVQWNLFF-ESPVEEDGENTLFTSISMILFDFFLXGWTWITIEAV-FPGYGIPIR 854
 Db 413 YMRREELNLMFNDSSLH--FSLGALVMMVLDIMMSIALVVDHRTSADSLRT 469
 QY 855 PWYPCPKSTVYFGESEDEKSHPG-----SNOKMSEICMEEP---TTL----- 895
 Db 470 LFDFE-----APEDEENQTDVTAQNTFRINQVRRVRRSMELQMPMASTSLNPNP 522
 QY 896 -----KLGVSIQNLVYKVRDGMKAVVDGLALNVEGQITSLGN 935
 Db 523 ADSLSLEGSTEADGADTARADIIYRNILKWTSTGEAVDGLSLRAVRCQCSILLGN 582
 QY 936 GAGTTTMSLITLFPPTSGTAVILGKDIRSEMSTIRONIGVCPHNVYFDMLVYEEHWM 995
 Db 583 GAGSTFSSIAIGIRPTNGRITICGYDVGNEGEGRRIIGCMCPQYINPLDQITVSEHLK 642

QY 996 FYARKLSEKHVKAEMQMALDVLGPPSKLSTQSGMORKLSVALAFVGGSKYVI 1055
 Db 643 LVYGLKAGREKDFQDMKRLSLDVKL--DFKENEAVALSGCMRKLCVCAALIGDSEVL 701
 QY 1056 LDEPTAGVDPYSRRIQWELLKTYRQRTIILSTHMDADVLGDRIAIISHGKLCVSS 1115
 Db 702 LDEPTAGMDPGARQDVOKLVEREKANRTIILTHYMDERLQDWFIIMSHGLVASGTN 761
 QY 1116 LFLNQDLQGYTLPLVKKKDVESLSLSCRNSSVYSVLKKKDSVSSSAGLGSHEST 1175
 Db 762 QYLKQKFGTGYLTV-----LDHNGDK 784
 QY 1176 LTIYSAISNLIRKHVSEALVEDIGHETVLYPYAAREGAFVELLFHEID----- 1226
 Db 785 RKMAV--ILTDVCHYHVEAERGEMHQOIEILIPARKKE--FVLPQALEAIQDNRYS 841
 QY 1227 -----DRISDGISSYISSETLEETLEIFLKAESSGVDAETSDGTLPARNRRAF 1275
 Db 842 NVFDNMENLTLSQALTELEMSTFSLNLTIEQVFTITGDG-----VKAIALSRNSRIS 894
 QY 1276 GDSQSCARPTEDDADPNDSIDIDPESRETDLSGMDGSGYQVGMKLTQOQFVALLMK 1335
 Db 895 HNSRNASEPLKPAQYDQTSSTKSADSYOK--LMDSQARGP-EKSGVAKMAQFISTMR 951
 QY 1336 RLIIARRSRKGFPAQVLPVAFVCIALVPSLIVPPFGKYPSLEDPMMYNQYTFVSND 1395
 Db 952 KFLYSRRNMAQLETVLIP--IILGLVSSL-----TLKSNNT 988
 QY 1396 PEDTGLLELNLKDPGCTGCMENPIPTPCQAGEEWTTAPVPTIMDLQNGNWT 1455
 Db 989 DQ-----FSVRSLS-----TP-SGIE-----PSKVVMRENKT-- 1014
 QY 1456 MGNPACQSSDKIKKMLPVCPPAGGLPPQPKQONADILQDLGTRNSDYLVKTYVO 1515
 Db 1015 -----IPEANERKILTRSGGEVLANVTNKKMLP 1044
 QY 1516 IIAKSLKAKKIWNERNRYGSGFLSVNTQALPPSQEVDNAIKMKHKLAKDSSADRLN 1575
 Db 1045 NTKLSL-----IGE-----MPPA----- 1057
 QY 1576 SLGRPMGTGIDTNNKVKWENKGMHAISFLVYNNALILANLQKGNPSHYITAENHP 1635
 Db 1058 TIGMTMNS-----DNEALFNNRYTHVLPITLSMTNRRLGTDAEISGVFLYSKSTN 1113
 QY 1636 LMLTQQLSEVALMTTSVDVLSICVIFAMSFYASVFLIOERYSKAKHLOFTGVP 1695
 Db 1114 SNLPSQL-----IDVLLAPMLILIFAMYSTFVMLIEERTCORAHQOFLTGISP 1164
 QY 1696 VTYMLSNFWMCMNVVYPATLVIIIFCPOOKSVSSTNLPVALLLLYGMSTPLMP 1755
 Db 1165 ITFYSALITGICITSLICILFLFELFLAF--HMWDHLATVILFWLFFSSVFET 1220
 QY 1756 ASEVFKIPSTAYVLTSVNLFTIGINGSVATFVLELFTDNKLNINDILKSVL-----IF 1810
 Db 1221 VSFLEQSRASAVLLIMQVVISGALLAVFLFM-----INRIDMELSLVNIEMELL 1275
 QY 1811 PHFLIGGLIDMKKNQAMALRGENRVSP-----LSMDLVGRNLMAVEGVFLI 1866
 Db 1276 PSYAFGSAIT-----TINTYG--MILPSEELMMMDHCGKNAMLMGTFVGSFAL 1322
 QY 1867 TVLIQYRFFIR-----PRPVNAKLSPLNDE--DEVDYRERORILLGGQNDILEIK 1915
 Db 1323 FVLLOEFKVFARFLSQVWTVVRSSHNNVQPMMDLPVCEVSESRERVRHVNSQNSALVI 1382
 QY 1916 ELTKIYRRKRPVADVICYGIPIPGCFGLLVNGAGKSTFEMKLTGDTTVTRGDAFLKN 1975
 Db 1383 DLTKTR--GFTAVNBLCLAVQDKCEFGLLGVNGAGKTTFTTNILLGQSFASGEMIGR 1440
 QY 1976 SILSNIEHYQNGYCPQDALTTELGTGHEHVEFPALLRGVPEKEVGVGMAIRKGLV 2035
 Db 1441 DVTELL-----SIGYCPQDALTMLDTLGTRESLEILLQMHGFENYKA--KALILCEVGM 1493

```

Oy 2036 KYGEKYACNIGSGKKRKSLSTMALIGEPYVFEIDPTGMDKARRFLMNCALSVKEGR 2095
      : : ||||| ||||| : ||| : ||||| ||||| ||||| : |||
Db 1494 AHADKLIVFYSGGGRKRKISGVALLAPQTOMITIDEPTAGIDPKARRREVWELLMCRESHN 1553
      : : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 2096 S-VYLTSHSMCECALCTGRMAIWNAGRFRCISYQHLKNRFGDYTTIVIRIANGSNPLDKP 2154
      | : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1554 SALMTSHSMCECALCTGRIRIIVNLKRGSLIAGSGSELKSLKGNVYTMTLISYEPNQRDMV 1613
      | : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 2155 VQDFEGLAFPGSVYIKERHRM---IQYLPSSLSL-ARITSILSSQKKRHIEDYSQ 2210
      || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1614 VQ-LVQTRLPNSVLTKTSTNTKLNLMQIIPKEKEDQWAKFEEMOALAKDGVKDFIIAQ 1672
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 2211 TTLDOVFVNF---KDQSDDDHLKDLDS 2234
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 1673 SSLEEFPLRLAGLDEDDLTSTVETS 1699
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

RESULT 12
S60124
transport protein homolog C4B84.4 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 13-Jan-1996 #sequence-revision 12-Apr-1996 #text-change 02-Feb-2001
C:Accession: S60124; S40724; S40725
R:Kershaw, J.
submitted to the EMBL Data Library, November 1995
A:Reference number: S60124
A:Accession: S60124
A:Molecule type: DNA
A:Residues: 1-1767 <KER>
A:Cross-References: EMBL:Z29117; NID:9439247; PID:91066912
C:Genetics:
A:Map position: III
A:Introns: 47/1: 112/3; 161/2; 220/2; 543/3; 574/3; 903/2; 1056/1; 1115/3; 1178/3; 1265/3;
C:Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homology
C:Keywords: ATP; duplication; nucleotide binding; P-loop; transmembrane protein
F:645-818/Domain: ATP-binding cassette homology <ABCI>
F:645-652/Region: nucleotide-binding motif A (P-loop)
F:764-769/Region: nucleotide-binding motif B
F:1457-1642/Domain: ATP-binding cassette homology <ABC2>
F:1474-1481/Region: nucleotide-binding motif A (P-loop)
F:1586-1591/Region: nucleotide-binding motif B

Query Match 13.0%; Score 1536.5; DB 2; Length 1767;
Best Local Similarity 23.7%; Pred. No. 1,5e-90;
Matches 522; Conservative 370; Mismatches 735; Indels 575; Gaps 66;

Oy 162 LSLPKSTVDKMLRADVILHVFLOGYQLHTSLCNGSKSEEMQLQ-GDOEVSSELGLPK 220
      : : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db 7 IPMHSTIVRLRLRYLQNEKAKRKAKESDIECSSSEMSNDSPRLRDGGLAESRIRPK 66
      : : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 221 KLAIAER-----VLNSN---MDLIKPLRLNSTSPPEKELAEATKTLIHSI 265
      : : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db 67 KMNRLRGESLLMKDWALLRRNKVWTLFELIIPCL-----L 103
      : : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 266 GTLQGEFLSMMSMDMQEVMFLNNVSSSSQIYQAVSRIVCGHPEGGLAKIKSLNWY 325
      : : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 104 GPLV-----VLVKNADHTSSPENIY----- 124
      : : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 326 EDNNYKALFEGNGTEDEAFETFDNSTPTPYCNDLKNLLESSPLSRIT--WKALPLVYGI 383
      || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 125 -DN-----FQYKGTVEY-----FLESNFIKPIYKWCILRSDVVVG-- 159
      || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 384 LYTPTPATROYMAEVNKTPOELAVFHDLEGMELSPKIWTFMNSQEMDLVRLDSR 443
      || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 160 -YTSKDAAKRIVDDLMKFAE-----RFQSAKIKLSYKNSSSE--QLLTIVL 204
      || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 444 DNDHFWEQQLDGLDWTADQIVAFLLAKHPEDYQSSNGSVYTWREAFNETNQAIPTISRFME 503
      || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 205 RND-----LPMLENFCAINS----- 220
      || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 504 CVNLKLEPIATEVWLINLKSMEILDERKFMAGIYFTGTPGSIELPHVKKYKIRINDIV 563
      || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 221 -----YAAGVVDEVDVYTKRL-----NRIILGLKTP 248
      || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

QY 1561 HLKLANDSSADRFPLNSLGRFMTGLDTRNNKVKWPNKGMHAISFLYINNAIIRANIOK 1620
 Db 1121 -----TIGTITMS-----DNLELFMRIRYHLEPTLISMIRALITGVDA 1161
 QY 1621 GENPSHYGTATRNHPLNLTQOLSEVALMTTSVDVLSICVIFAMSEFPASVFLIOER 1680
 Db 1162 EISSGVFLSKSTSNLSNLLPSOL-----IDVLLAPMLILIFAMVTSFVFLIEER 1212
 QY 1681 VSKAKHLOFISGVKPVYIWLNSFWMDMCNYYVPATVLIITICQOQSYSSTLPLVAL 1740
 Db 1213 TCGFNAHQGLTGISPTTFYSALITDGLIYSLICITLFLFPLAF-----HMYDHLATVIL 1268
 QY 1741 LLLIGWSTPLMPYASVFKIPSTAVYVLTSVNLFIGINSVATFVLETPDKLNIN 1800
 Db 1269 FWFLEFFSSVPFIYASVFLQSPSKANVLLITVOVYISGAALLAVFLIM-----INID 1323
 QY 1801 DIKSVFL-----IPPHCLGRGLDMKKNAMDALEERGENFVSP-----LSWDLVGR 1851
 Db 1324 EMLKSLIVNIIMFLLPSTAFSAIT-----TINTYG---MILPSEELMNMWDHCK 1370
 QY 1852 NLFAMAVEGVVEFLLTVDLQYRFFIR-----PRPVNAKLSPLNDE---DEDVRRRQ 1900
 Db 1371 NAMLMGTFCVSFALFVLLQFKFVRRLSQYVTVRRSHNNVQPMGDDLPVCSVSEBER 1430
 QY 1901 RLDOGGONDILLETIKELTKIYRKRKPAVDRIQVIRPGECFGLIGVNGAKSSTFKMLT 1960
 Db 1431 RHRVNSONSALVTKDLTKF--GRFPAVNEELCLAVDQKCFGLIGVNGAKTTFITNLT 1488
 QY 1961 GDTTVTRDADFLNKNRSLISNIEVHOMWGYCOPDATTETLLTGREHVEFPALIRGPEKE 2020
 Db 1489 GOSFASSEAMIGRDVTELL-----SITGPOPDALMDLTRESLEIILAOHNGENYK 1543
 QY 2021 VGKGEAMIRKGLGVKGEKYAGNYSGGNRKSLTAMALIGRPVFLDEPTGMDPKAR 2080
 Db 1544 A--KAELLIECVGMIAHADKLIVREYSGGOKRKISGVALLAPQMLIDEPFAGIDPKAR 1601
 QY 2081 RFLNMCALSVYKEGRS--VLTSHSMCEPCALCTRMALMVNGRCRIGSYQHLKNRPGDY 2139
 Db 1602 REVWELLKMCREHNSALMTLSHMDCEALCSRIAVLNKSLIALIGSSQELSLATGNV 1661
 QY 2140 TIVVRINGSNPDLKPVODFFGLAFPGSVLKEKHRNM---LOYOLPSSSL--ARIFSTLS 2195
 Db 1662 TMLSLTEPQORDMVYO--LVQTRLPNSVLTSTNKTMLKMOIPEKEDCSAFENYO 1720
 QY 2196 OSKRRLHIEDYSVOTLDOVFNFNA---KDQSDDLKLDLS 2234
 Db 1721 ALAKDLGVKDFITLAQSSLETFURLAGLDDLDLHSTIVEIS 1762

RESULT 13

T00826

hypothetical protein T3266.22 - Arabidopsis thaliana (fragment)

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 12-Feb-1999 #sequence: revision 12-Feb-1999 #text: change 23-Mar-2001

C:Accession: T00826

R:Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kaut

submitted to the EMBL Data Library, November 1997

A:Description: Arabidopsis thaliana chromosome II BAC T3266 genomic sequence.

A:Reference number: 214163

A:Accession: T00826

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1246 <R0U>

A:Cross-References: EMBL:AC002510; NID:g2618683; PID:g2618705

A:Experimental source: cultivar Columbia

C:Genetics:

A:Map position: 2

A:Intons: 33/3; 95/2; 113/3; 137/3; 168/3; 361/3; 421/2; 432/3; 493/3; 521/3; 535/3; 56

A:Note: T3266.22

Query Match 11.9%; Score 1402; DB 2; Length 1246;
 Best Local Similarity 28.7%; Pred. No. 4.3e-82;
 Matches 415; Conservative 195; Mismatches 388; Indels 450; Gaps 43;

QY 1019 VGLPSSKTKSKTSOLSGGMORKLSALAFVGGSKVVLIDETAGVDPYRRGIGWELLKY 1078
 Db 1 VGL-SDKINTVLVRLSGCMKRLSLGIALIGNSKVILLDEPTSGMDYSRMLTMOILIKKI 59
 QY 1079 RGRFTTILSTHMDADVLGRIATISHGKLCCYSSSLFELKNOLGRTGYTLTKRVKVES 1138
 Db 60 KGRITILLTTHSHMDEAFELGDRIGIMANGSLKCCGSSIFLKHNGVGYTLTLK----- 113
 QY 1139 LSSCNSSSTVYLKKEISVSQSSDAGLSDHSDTLTIDVSAISNLIRKHVSEARLVE 1198
 Db 114 -----TSPY-----VSAAHYIHRHIPSATCVS 136
 QY 1199 DIGHELTVLPYEAARKGAFVLEFHEIDRL-----SDP-GISSYGISETTLEFIPKVE 1253
 Db 137 EYGNISFRLP--LASLPCEFMRELESCKMSKSDYRPTQSTGISTVTLLEVFLRAV- 193
 QY 1254 ESGVDAEISDGTLPARRNRRAFGDKQCLRPFTEDDAADPNDSIDIPESRETDLSGMDG 1313
 Db 194 GCNLDLE-----DKQ-----EDIFVSPDTKSSLYIGISNQ 223
 QY 1314 KGSYQVK-----GKRLTQ-----QGFVALLM----- 1334
 Db 224 KSMQPKLLASCNDAAGVITTSYAKAFRLIVAAVWTLIGFISIQCCGSIISRSMEFWRHC 283
 QY 1335 -----KRLIARRSRKGFQAQIVLPAVFCIALVFLVPPGKYPSLELOPMWYTBQYT 1389
 Db 284 KALFIRKARSACRDKRTVAFOFIIPAVFLFLGLF-----LQLR----- 323
 QY 1390 FVSNDAPEDEGTLELLNALKKDGFPGTRCMEGRPITDPTCQAGEBWTTAPVQTIMLF 1449
 Db 324 --HPDOKSTILTAYFNPLISGKG-----GGGPIPFD-----LSVPAKEVAQYI 366
 QY 1450 QNG-----NWTQNPSPACOCSSDKIKKMLPVCPAGAGLPPQRKONTDILQDLGR 1503
 Db 367 EGGWIOPLRNTSYKFPNP-----KALADAI--DAAGP 397
 QY 1504 NISDYLKTYVQVLIASLKNKTIWNEFRYGGSLGVSNTQALPSPQEVNDAIKOMKHLK 1563
 Db 398 TLQPTLS-----MSEFLMSSPD-----QSYOSSRE-----G 424
 QY 1564 LARDSSADRFPLNSLGRFMTGLDTRNNKVKWPNKGMHAISFLYINNAIIRANLOGEN 1623
 Db 425 LSHSDCNHPDGLG-----YVLLHNGTQOHGPIYINMRAILR--LATGN- 470
 QY 1624 PSHYGTATRNHPLNLTQ--QOLSEVALMTTSVDVLSICVIFAMSEFPASVFLIOERVS 1682
 Db 471 -KMTIOTRNHPLPPTKTQRIORHDLDAFSAALIYNI---AFSFIAPSAVPIVKEREV 525
 QY 1683 KAKHLOFISGVKPVYIWLNSFWMDMCNYYVPATVLIITICQOQKSYSSSTNLPLVALL 1742
 Db 526 KAKHQQLISGVSVLSTWISTYWDISFLPSTRAILLIFVAFGLEOFTIGIGRPLPVLM 585
 QY 1743 LLYGWSITPLMPYASVFKIPSTAVYVLTSVNLFI-----GINGS 1782
 Db 586 LBYGLIASSTYCLIPFFTEHSMQAOTSSYVLLPISLFWSPSSNVLWVHFFSGILM 645
 QY 1783 VATEVLELF--VDNKLNNINIDI-----LKSVELIPHFCLGRLIDM--VKNOAMADALE 1833
 Db 646 VLSFVWGLIPATASANSYLKLELILFRYALQNFRLSPGFESDGLASTALROGMKDKS 705
 QY 1834 RGENREVSPLSWDLVGNLFAMAVEGVFLLTVDLQYRFFIRPRVNAKLS----- 1886
 Db 706 H-----GVFEMVNTGASICYGLVLRLEYCRSMILTSF--HGIDIKSLIYITIGA 755
 QY 1887 -----PL-----NDEEDVRRBORILLDGGQNDILEIKE 1916
 Db 756 SLEFELIDRYVYSSTFTEPLKSDTAISDMEDDIDVOEDRVDYISGLSDMTMYLQGN 815
 QY 1917 LTKIYRRKR-----KPAVDRIQVIRPGECFGLIGVNGAKSSPFKMLTGDITYTRDADL 1972
 Db 816 LRKVPDGMKHHGRVAVAVQSLTFSVQAGECGFGLTGNAGKTTTLMISGEPTTSGTAFI 875

```

QY 1973 NKNLSLNHEVHÖNNQNGCOPOPATIELLTGHEHVEFPALLRGVEKPEVKGEMWALRKL 2032
Db 876 FGKDIVASPAIRÖHIGYCPÖPALFEYLTVEKHELTARIKGVNDHRIDNVYTERKLEVF 935
QY 2033 GLVYKEGYVAGNYSGNKRRKLTAMALLIGPVPVLEDEPTTGMDPKARFLMN--CALSV 2090
Db 936 DLKHSIKRPEFTLSTGNGKRLSVALIMIDPPIVLTDEPSTGMDPVARFMMVDYISRLST 995
QY 2091 VKEGRSVYLTSHSMECECALCTPMALMNGRPRCLGSOHLKNRPG----- 2136
Db 996 RSGKTAIVLTTHSHMNEAOALCTRGIMVGGRIJGSPDHLKTRGNHLEVEFPYNGVK 1055
QY 2137 -----DGYV-----I 2141
Db 1056 PNEYSVNLENEFCOITIOQWLFNVPYTPQPSRLLDGLEYCIGVSDSITPPTASASEISLSPEM 1115
QY 2142 VVRIA-----GSNP-----DLKPV 2155
Db 1116 VQRIAKFLGMDQVRSTVYPPLEPEYDRFPDQJLSEQLFRGGIGPLRILFAPMWLTKEKFSAL 1175
QY 2156 QÖDFEGFLAPGSVLKEKHRNMLQYÖLPSSLS--SLARIESILSOGKRLHIEDYSVSQÖTL 2213
Db 1176 DSFIOSSEPGATFEKSCNGLSIKYÖLPFEGGGLSLADAFGHLERNNRNLRIGLAEYSISQÖTL 1235
QY 2214 DQVFVNFA 2221
Db 1236 ETIENHFA 1243

```

RESULT 14
 T27121
 hypothetical protein Y53C10A.9 - *Caenorhabditis elegans*
 C:Species: *Caenorhabditis elegans*
 C:Date: 15-Oct-1999 #sequence.Revision 15-Oct-1999 #text.Change 04-Mar-2000
 C:Accession: T27121
 R:White, S.
 submitted to the EMBL Data Library, November 1998
 A:Reference number: Z20314
 A:Accession: T27121
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1564 <MIL>
 A:Cross-references: EMBL:AL033536; PIDN:CAA2142.1; CESP:Y53C10A.9
 A:Experimental source: clone Y53C10A
 C:Genetics:
 A:Gene: CESP:Y53C10A.9
 A:Introns: 43/3; 92/2; 148/2; 226/3; 354/1; 712/3; 817/1; 875/1; 916/3; 984/3; 1069/2; 1
 A:Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homology

[illegible]

QY	723	SVFAVTVILLOCFILSTLFSKANAANAAAGGIIITFLYLYLVCA-----MODY-----GF	773
Db	278	FMYTCHCVSFSLCTSLTPGKRIYEGWVIWITLILAMHLSLEBEDMFLRWFPLNPT	337
QY	774	TUKTFASLSPVAFVGCCEFAFLEBEOGIGVOMDNI,FSAPVEEDGFNLJTSISMPLDPT	833
Db	338	SLKFLVATLTLASPGNPTPTSLAFSSK-----KKTQSAAYFGIMISCVLVAIAIF	391
QY	834	LYGVMWYIEAVPPGOYGIIPRWYFECTSYN--FGESDEKS---HFGSNOKRMSETCM	888
Db	392	MEKLYTFLVGHAF-----KRFWRILFGSKGKSKTEERGDEVERSTILO	436
QY	889	EEEPHLHK-----LGVSIONLVKYYRDMKVAAVDGALNFEQOITSEFGHAGAKTTMS	944
Db	437	CKEIVEGRGSLADIETLSGLVKYQNGEK-AVANGLSLARIQOVSILLGHGCKSKTFFG	495
QY	945	ILVGLFPPTSGTAVILGKDIRSEMSTTRONLGVCPDOHNVLFDMLTVEEHIFVRLKGLS	1004
Db	496	MITGHQATEBKWVIGGIDNANRARAARELIGCPOYPIYDELTVWHLRLVALVAGRS	555
QY	1005	-EKHHKAMEQMDALDVLPSKLSKSTSOLOGSMOKRISVALAFVGGSKVILDEPAGV	1063
Db	556	GGSDFKMAESLKQIEL-TDKRNTLAKNLSGOKRKLKVCMAAMIGGSRVILDLDEPTAG	614
QY	1064	DPSYKRGIMELLLKYRGRTIILSTHHMEDADVLDRIATISHGKLCVGSSEFLFKMQLG	1123
Db	615	DPSAIDVQNNMLVYKADRTILTLTHYMDAEKLGDMIFVMSHGKMAAGSKHYLKQYGG	674
QY	1124	TGYIYTLV-----KKDVSSLSLSCRNSSYIYLAKDVSQSSSDAGLGSDES	1173
Db	675	GGMLTLVFKSHDPMRPRKSYETAVDYCKYCYSTA--LVKKE-----GOMIEI	722
QY	1174	DLTLTDVAISNLRKRKVAERARVEDIGHETLVLYPEAKGAEVLEFHEIDRLSDLG	1233
Db	723	SLLEEKRLPTLTK--ILSSVMEEDYNNP-----EPQALEPIDQEKCPLE	767
QY	1234	ISSVGETSTLLEELFLKVAEE-----SCVDAETSDGTLPARNRKRAFGDKOSTLRPT	1286
Db	768	IATIGVSSMSLEOVFIKIGDECDDINGGCVDKRT-----ERQEKFSTLV	812
QY	1287	EDDADPDDSDIDDESEBETDILSGMDGKGSYOYKCMKLTKQOQFVALMLKRLIARSRKGG	1346
Db	813	OYKIDOPK-----QGRSKLMYVMVALLQKRAVILARNVQ	847
QY	1347	FFAQIVLPAVFAVCTALVPSLIVPEPGKYPSELDOPMYWEDQTEVNSDAPEDTGTLELN	1406
Db	848	ITLOJLPL-----LTLTLMFAVBFLL--LEBKPRKSLDISFDPISOYPRHSTVLLDLEN	898
QY	1407	ALTROPGRGTGCMGPNLIPDTPROQAGEEMTTAPVQITIMDLFQNGNMTMQNPSPACCS	1466
Db	899	E-----N	900
QY	1467	SDRIKMLPYCPGAGGLPPOROKONTADILDLTGRNISDYLVKTYVOIILAKSLKNTW	1526
Db	901	DORLANVYL-----NFSNLF-----EYVFKTL	921
QY	1527	VNEFRYGGFSLGYSMTQALPPSOEVNDAITQMKHKLIAK--OSSADRPILNSLGRM----	1581
Db	922	-----GFIYKVNK-----KDSKPYKISOGDNNAILMNTIATASMYLRD	960
QY	1582	-----TGCLDRNNVKKVFNKKGHAISSEFLVNIINNAILRANLQKENSHTGIFANHPNL	1638
Db	961	PSVTKLPHYTSNVIIMND-----PKIKYEGLASF-----	989
QY	1639	TKQOULSEVALMTTSADVLSICVIFAMSEVPASVYVFLIOERKSKAKHLQFIOSYKPVLY	1698
Db	990	-----FLENFIFPLVLAGIFIOS-----TYVLLIEKICKPAHOQOYLTGSTIAY	1034
QY	1699	WLSNFWVMQCNVVPATVILVILIFCFOOKSVYSTNLP--VIALLLLLG--WSITPIY	1754
Db	1035	WGVVELMDF-----LFTFPLLLTYTGFLISGVLOGHIHEITVIFPGLGLFYAPRLVY	1086
QY	1755	PASFEFKIPSTAYVVLVSVNLFIGINGSVATFVLELEFOTDKLNNINDILKSVEFLIF-BHF	1813

```

Db 1087 LNSALNTPTRGNFL----YMFCCIPWLAYSIVSELDHNPPIOKTSDLETGFRIFNPSI 1143
QY 1814 CIGRGIDLVKNOAMADALEREENREVSPLSWDLVRNLFAMAVEGVFFL----- 1865
Db 1144 GFLAGIMKTAALNYPKSGIDKHFEHL-----TNLM--TYEGIEFELMFLFFGCI 1190
QY 1866 -ITVIT-----QYRFFI-----RPRYNKLSPLNDEDEVRERORILLDGGQNDIL 1912
Db 1191 FTILIGCATLPPFRACRGTRRRSOPERRRYKGLIESCAVKEEDLVQEVDNKEIVL 1250
QY 1913 EIKELTKIYRRKRPAYDRIQVIGPPGEGFLIGVNGAKSSTFKMLTGDITVTRGDAFL 1972
Db 1251 VIDGLVKGDKGFR--AVNDLSISVGHGECFGLGANGAGKTTTFTDITGLTPTGSAFI 1308
QY 1973 NKNLSILNHEVHOMGCPQPDATTELLTGREHVEFPALLRGVPEKVGKGEWAIRKL 2032
Db 1309 DGHDTTETIH---IGYCPQPDAMLQOISCRQTLRIMAKLQGYF--NVKEVELVLDV 1361
QY 2033 GLVYGEKYAGNYSGKNRKISTAMALIGPPVYVFLDEPTGMDPKARRFLMNCALSYK 2092
Db 1362 GMSDFGYKLVKNCSSGQKRIKISGIALMSRATCITLDEPTAGIDPRARRIWDITHEMR 1421
QY 2093 EGR-SVILTSHSMECEALCTRMALMVNGRFRCLGSVOHLKNRFGDGYTIIVRIAGSNPD 2151
Db 1422 QAKCSIVLTSHECEALCTRIGILRKEMIALGTSQSLKSQYNTYMWTL-ILNSLED 1480
QY 2152 LKPVODFFGLAFPGSVLKEKHNM---LOYOLPSSL-SLARIFSILSOSKKRLHIEDYS 2207
Db 1481 LESVCVIVSEKMPDAVLTPSSSLTTSIVWELPKRSIDKSEKYNQVEVLAKKANAKDYM 1540
QY 2208 VSQTLTDOVFVNFPAKQSDDD 2228
Db 1541 LTQASLEDPTFRLITTEEEEE 1561

```

RESULT 15

T47150

hypothetical protein DKFZp547P193.1 - human (fragment)

C:Species: Homo sapiens (man)

C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 02-Sep-2000

C:Accession: T47150

R:Blocker, H.; Boecker, M.; Brandt, P.; Mewes, H.W.; Well, B.; Wiemann, S.

submitted to the Protein Sequence Database, March 2000

A:Reference number: 224376

A:Accession: T47150

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-373 <AAA>

A:Cross-references: EMBL:AL162060

A:Experimental source: fetal brain; clone DKFZp547P193

A:Note: DKFZp547P193.1

C:Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homology

Query Match 8.68; Score 1013.5; DB 2; Length 373;

Best Local Similarity 65.0%; Pred. No. 8.4e-58;

Matches 195; Conservative 40; Mismatches 64; Indels 1; Gaps 1;

```

QY 1928 AVDRICVIGPPGEGFLIGVNGAKSSTFKMLTGDITVTRGDAFLMNCALSYK 1987
Db 7 AVDRICVIGPPGEGFLIGVNGAKSSTFKMLTGDITVTRGDAFLMNCALSYK 1987
QY 1988 MGYPQPDATTELLTGREHVEFPALLRGVPEKVGKGEWAIRKLGLVYGEKYAGNYSG 2047
Db 67 LGYCPQCDALDFDLTAHREHQLYTRLRGISMKDEARVYKMALEKLELTKYADKPAGIYSG 126
QY 2048 GNRKRLSTAMALIGPPVYVFLDEPTGMDPKARRFLMNCALSYKSGSVVLTSHSMEEC 2107
Db 127 GNRKRLSTAMALIGPPVYVFLDEPTGMDPKARRFLMNCALSYKSGSVVLTSHSMEEC 186
QY 2108 EALCTRLAIVNGRRLCLSGIOHLKNRFGDGYTIIVRIAGSNPDLPKQVDFGLAFPGSV 2167

```

```

Db 187 EALCTRLAIVNGRRLCLSGIOHLKNRFGDGYTIIVRIAGSNPDLPKQVDFGLAFPGSV 245
QY 2168 LKEKRRMLOVOLPSSLSLARIFSILSOSKKRLHIEDYSVSQTLTDOVFVNFPAKQSDDD 2227
Db 246 LKEKRRMLOVOLPSSLSLARIFSILSOSKKRLHIEDYSVSQTLTDOVFVNFPAKQSDDD 305

```

Search completed: March 7, 2003, 08:51:22
 Job time: 121 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 7, 2003, 08:40:59 ; Search time 86 Seconds
(Without alignments)
3503.252 Million cell updates/sec

Title: US-09-595-526C-2
Perfect score: 11797
Sequence: 1 MACMPCQLLLMKNLTFRRR.....VDVAVLTSPLQDEKVESYV 2261

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 908470 seqs, 133250620 residues
Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: A-Geneseq_101002:*
2: /SIDS2/gcgdata/geneseq/geneseqp-emb/AA1980.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqp-emb/AA1981.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqp-emb/AA1982.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqp-emb/AA1983.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqp-emb/AA1984.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqp-emb/AA1985.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqp-emb/AA1986.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqp-emb/AA1987.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqp-emb/AA1988.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqp-emb/AA1989.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqp-emb/AA1990.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqp-emb/AA1991.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqp-emb/AA1992.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqp-emb/AA1993.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqp-emb/AA1994.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqp-emb/AA1995.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqp-emb/AA1996.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqp-emb/AA1997.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqp-emb/AA1998.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqp-emb/AA1999.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqp-emb/AA2000.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqp-emb/AA2001.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqp-emb/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	11797	100.0	2261	22 AAB31361	Amino acid sequence
2	11797	100.0	2261	22 AAB31365	Amino acid sequence
3	11793	100.0	2261	21 AAB38117	Human ABC1 cholest
4	11793	100.0	2261	22 AAB31362	Amino acid sequence
5	11792	100.0	2261	22 AAB31366	Amino acid sequence
6	11792	100.0	2261	21 AAB38109	Human ABC1 cholest
7	11789	99.9	2261	21 AAB38082	Human ABC1 cholest
8	11789	99.9	2261	22 AAB71749	Human ABC1 protein
9	11789	99.9	2261	22 AAB31363	Amino acid sequence
10	11789	99.9	2261	22 AAB31367	Amino acid sequence

11	11786	99.9	2261	21 AAB38111	Human ABC1 cholest
12	11786	99.9	2261	21 AAB38114	Human ABC1 cholest
13	11786	99.9	2261	21 AAB38115	Human ABC1 cholest
14	11785	99.9	2261	21 AAB38105	Human ABC1 cholest
15	11785	99.9	2261	21 AAB38110	Human ABC1 cholest
16	11784	99.9	2261	21 AAB38113	Human ABC1 cholest
17	11784	99.9	2261	21 AAB38116	Human ABC1 cholest
18	11783	99.9	2261	21 AAB38112	Human ABC1 cholest
19	11777	99.8	2261	21 AAB38104	Human ABC1 cholest
20	11774.5	99.8	2260	21 AAB38106	Human ABC1 cholest
21	11771	99.8	2261	21 AAB38107	Human ABC1 cholest
22	11767	99.7	2259	21 AAB38107	Human ABC1 cholest
23	11767	99.7	2261	21 AAB38102	Human ABC1 cholest
24	11767	99.7	2261	21 AAB38102	Human ABC1 cholest
25	11767	99.7	2261	21 AAB38102	Human ABC1 cholest
26	11767	99.7	2261	21 AAB38102	Human ABC1 cholest
27	11767	99.7	2261	21 AAB38102	Human ABC1 cholest
28	11764	99.7	2261	21 AAB38102	Human ABC1 cholest
29	11763	99.7	2261	21 AAB38102	Human ABC1 cholest
30	11761	99.7	2261	21 AAB38102	Human ABC1 cholest
31	11760	99.7	2261	21 AAB38102	Human ABC1 cholest
32	11760	99.7	2261	21 AAB38102	Human ABC1 cholest
33	11760	99.7	2261	21 AAB38102	Human ABC1 cholest
34	11759	99.7	2261	21 AAB38102	Human ABC1 cholest
35	11759	99.7	2261	21 AAB38102	Human ABC1 cholest
36	11737	99.5	2261	21 AAB38102	Human ABC1 cholest
37	11433	96.9	2201	21 AAB38102	Human ABC1 cholest
38	11433	96.9	2201	21 AAB38102	Human ABC1 cholest
39	11433	96.9	2201	21 AAB38102	Human ABC1 cholest
40	11433	96.9	2201	21 AAB38102	Human ABC1 cholest
41	11015	93.4	2130	22 AAB38108	Human ABC1 cholest
42	11015	93.4	2130	22 AAB38108	Human ABC1 cholest
43	7988	67.7	1525	22 AAB38108	Human ABC1 cholest
44	5858.5	49.6	1144	22 AAB38108	Human ABC1 cholest
45	5849	49.6	1144	22 AAB38108	Human ABC1 cholest

ALIGNMENTS

RESULT 1	AAAB31361	standard; Protein: 2261 AA.
ID	AAAB31361	
XX	AAAB31361	
AC	AAAB31361	
XX	AAAB31361	
DT	20-APR-2001	(first entry)
DE	Amino acid sequence of a human ABC1 polypeptide.	
XX	Human; adenosine triphosphate binding cassette protein 1; ABC1;	
KW	apolipoprotein-mediated mobilization; cholesterol; Tangier disease;	
KW	chromosome 9q22-q31; heart disease; hypercholesterolemia;	
KW	atherosclerosis; cholesterol transport.	
OS	Homo sapiens.	
XX	WO200078972-A2.	
PN	28-DEC-2000.	
XX	16-JUN-2000; 2000WO-US16765.	
PF	18-JUN-1999; 99US-0140264.	
XX	PR 14-SEP-1999; 99US-0135872.	
PR	PR 19-NOV-1999; 99US-016573.	
XX	(CVTH-) CV THERAPEUTICS INC.	
PA	Lawn RM, Wade D, Garvin M;	
XX	WPI; 2001-137812/14.	
DR		
XX		

PT Adenosine triphosphate (ATP) binding cassette (ABC) polynucleotide,
 PT useful for the development of agents for the treatment of heart disease
 PT and other disorders associated with hypercholesterolemia and
 PT atherosclerosis -
 XX
 PS Disclosure: Page 128-143; 215pp; English.
 XX
 CC The present sequence represents a human adenosine triphosphate (ATP)
 CC binding cassette protein (ABC) 1 polypeptide. ABC1 resides in cell
 CC membranes and utilizes ATP hydrolysis to transport a wide variety of
 CC substrates across the plasma membrane. ABC1 is a pivotal protein in
 CC the apolipoprotein-mediated mobilisation of intracellular cholesterol.
 CC ABC1 is defective in Tangier disease, a genetic disorder
 CC characterised by abnormal HDL-cholesterol metabolism. The ABC1 gene is
 CC localised to chromosome 9q22-9q31. The ABC1 genes and proteins are
 CC useful for developing pharmaceutical agents for the treatment of heart
 CC disease and other disorders associated with hypercholesterolemia and
 CC atherosclerosis. The genes are useful for developing screening assays to
 CC screen for compounds that regulate the expression of genes associated
 CC with cholesterol transport. The genes and proteins are also useful for
 CC are also useful as diagnostic indicators of cardiovascular disease and
 CC other disorders associated with hypercholesterolemia.
 CC
 XX
 SQ Sequence 2261 AA:

Query Match 100.0%; Score 11797; DB 22; Length 2261;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MACMFOLELLMKNLTFRRROTCOLLEVAMPLEFIIILISVLSLTPPEOHECFPNKA 60
 DB 1 MACMFOLELLMKNLTFRRROTCOLLEVAMPLEFIIILISVLSLTPPEOHECFPNKA 60
 QY 61 MPSAGTLPWVGITICANNPCFRPTPGEPAGVGVGNFNKSTIVARLPSDARLLISQKPT 120
 DB 61 MPSAGTLPWVGITICANNPCFRPTPGEPAGVGVGNFNKSTIVARLPSDARLLISQKPT 120
 QY 121 SMKMRVYLRTLOOIKKSSNKLQDFLVNENFSGFLYHNLSLPKSTYDKMLRADVILH 180
 DB 121 SMKMRVYLRTLOOIKKSSNKLQDFLVNENFSGFLYHNLSLPKSTYDKMLRADVILH 180
 QY 181 KVFIOGVLHLTSLCNGSKSEMIQLGDQVSELGLPREKLAARVLRKSMIDLKPL 240
 DB 181 KVFIOGVLHLTSLCNGSKSEMIQLGDQVSELGLPREKLAARVLRKSMIDLKPL 240
 QY 241 RTLNSTSPFSKELAEATKTLHSLGTLAQELFSSMSKSDMOEVLTVNSSSSSTQI 300
 DB 241 RTLNSTSPFSKELAEATKTLHSLGTLAQELFSSMSKSDMOEVLTVNSSSSSTQI 300
 QY 301 YQAVSRIVCGHPGGGLIKSLNMYEDNNYKALFGNGTEEDAEFFYNSTTPYCNOLMK 360
 DB 301 YQAVSRIVCGHPGGGLIKSLNMYEDNNYKALFGNGTEEDAEFFYNSTTPYCNOLMK 360
 QY 361 NESSPSLRITMKALKELLVGLILYTPPTPATROYMAEVNKTFOELAVFHDLEGMEELS 420
 DB 361 NESSPSLRITMKALKELLVGLILYTPPTPATROYMAEVNKTFOELAVFHDLEGMEELS 420
 QY 421 PKIWTMENSOEMLVRLMLDSRDNDHFEQOLDLMTADIDYAFIAKHEEDVOSSNGS 480
 DB 421 PKIWTMENSOEMLVRLMLDSRDNDHFEQOLDLMTADIDYAFIAKHEEDVOSSNGS 480
 QY 481 YTWRAEFNEIQAIRITISRFMECVNLKLEPIATEVWLINKSMELLDERKFMAGIYFTG 540
 DB 481 YTWRAEFNEIQAIRITISRFMECVNLKLEPIATEVWLINKSMELLDERKFMAGIYFTG 540
 QY 541 ITPGSIELFHNHYKIRMDINDVERTNKKDGYMDPGRADPFEDMRYVWGGEAYLODV 600
 DB 541 ITPGSIELFHNHYKIRMDINDVERTNKKDGYMDPGRADPFEDMRYVWGGEAYLODV 600
 QY 601 EOAIIIRVLTEKKTGYVMQMPYCYVDLIFLRVMSRSMPLFMTLAWISVAVIIGIY 660
 DB 601 EOAIIIRVLTEKKTGYVMQMPYCYVDLIFLRVMSRSMPLFMTLAWISVAVIIGIY 660

QY 661 YEKEARLKEFTMRIMGLDMSILMFNFISSLLPLVSAQLVLILKGNLPPYSDPSVAFV 720
 DB 661 YEKEARLKEFTMRIMGLDMSILMFNFISSLLPLVSAQLVLILKGNLPPYSDPSVAFV 720
 QY 721 FLTSVAVVTLIOCFLLISTFLFSRANLAAAGGIIYFTYLPVLVLCVAMQDVGFTLKIFNS 780
 DB 721 FLTSVAVVTLIOCFLLISTFLFSRANLAAAGGIIYFTYLPVLVLCVAMQDVGFTLKIFNS 780
 QY 781 LLSPAFPGCEYFALFEEOGIGVQMDNLESPEVEDGFMULTSISMPLDFEFLYGVMTW 840
 DB 781 LLSPAFPGCEYFALFEEOGIGVQMDNLESPEVEDGFMULTSISMPLDFEFLYGVMTW 840
 QY 841 YTEAVFPGOYGIIPRWYFPCRSYWFGESEDEKSHPSNCKRMSELCMEEPFHLKLVGS 900
 DB 841 YTEAVFPGOYGIIPRWYFPCRSYWFGESEDEKSHPSNCKRMSELCMEEPFHLKLVGS 900
 QY 901 IONLVKVRDGMKVAVDGLALNFYEGQITSFLGHNGAGKTTMSILTGLFPPTSGTAYTL 960
 DB 901 IONLVKVRDGMKVAVDGLALNFYEGQITSFLGHNGAGKTTMSILTGLFPPTSGTAYTL 960
 QY 961 GKDIRSEKSTIROMLVGCPQHNVLFQMLTYEEHIMFYARLKGISEKHVAAEEMOALDVG 1020
 DB 961 GKDIRSEKSTIROMLVGCPQHNVLFQMLTYEEHIMFYARLKGISEKHVAAEEMOALDVG 1020
 QY 1021 LPSSKLKSKTSQSLSGNQRRLSVALAFVGGSKVVIIDEPTAGADPYSRGIMELLKTYRO 1080
 DB 1021 LPSSKLKSKTSQSLSGNQRRLSVALAFVGGSKVVIIDEPTAGADPYSRGIMELLKTYRO 1080
 QY 1081 GRTIILSTHNMDEADVGLDGLIATIIISHGKLCVSSSLFLKNQIGTYITLVKKDVESSL 1140
 DB 1081 GRTIILSTHNMDEADVGLDGLIATIIISHGKLCVSSSLFLKNQIGTYITLVKKDVESSL 1140
 QY 1141 SCRNSSSTVSYLKKEDSVSSSDAGLGSDBHSDTLTDIVASISMLIRKHSEARLVADI 1200
 DB 1141 SCRNSSSTVSYLKKEDSVSSSDAGLGSDBHSDTLTDIVASISMLIRKHSEARLVADI 1200
 QY 1201 GHETLYVLPPEAKKEGAFELFHEIDDRISDGLISYGISETTLEIFLKYAESGVDAE 1260
 DB 1201 GHETLYVLPPEAKKEGAFELFHEIDDRISDGLISYGISETTLEIFLKYAESGVDAE 1260
 QY 1261 TSGCTLPARRNRRAFQKOSCLRPTEDDAADPNOSIDIPESRETDLSGMDGGSYOVK 1320
 DB 1261 TSGCTLPARRNRRAFQKOSCLRPTEDDAADPNOSIDIPESRETDLSGMDGGSYOVK 1320
 QY 1321 GMRLTQOQVALLMKRLILARSKRGFPAQIYLPVAVYICALVPSLVPFGKXPSLEQ 1380
 DB 1321 GMRLTQOQVALLMKRLILARSKRGFPAQIYLPVAVYICALVPSLVPFGKXPSLEQ 1380
 QY 1381 PMWYNQOYTFVSNDAPEDTGTLELINALTKDPGFCTRCMGPNIPDPPOAGGEEMWTAP 1440
 DB 1381 PMWYNQOYTFVSNDAPEDTGTLELINALTKDPGFCTRCMGPNIPDPPOAGGEEMWTAP 1440
 QY 1441 VQOTIMDLFONONNMWONPSPACOCSSDKIKKMLPVCPCGAGGLPPOROKONTADILQDL 1500
 DB 1441 VQOTIMDLFONONNMWONPSPACOCSSDKIKKMLPVCPCGAGGLPPOROKONTADILQDL 1500
 QY 1501 TGRNISDYLKVVYVOIILKSLKKNKIWNEFRYGGFSIGVNSTQALPPSOVNDATQOMK 1560
 DB 1501 TGRNISDYLKVVYVOIILKSLKKNKIWNEFRYGGFSIGVNSTQALPPSOVNDATQOMK 1560
 QY 1561 HLKLAQSSADRFNLNSLGRFMTGLDTRNNVKWPNFNKGWHAISPLVNIINNALIRANLQK 1620
 DB 1561 HLKLAQSSADRFNLNSLGRFMTGLDTRNNVKWPNFNKGWHAISPLVNIINNALIRANLQK 1620
 QY 1621 GENPSHYGITAIFNPNLNTKQOLSEVALMTTSVDVLSICVITFAMSFVPASVFLIOER 1680
 DB 1621 GENPSHYGITAIFNPNLNTKQOLSEVALMTTSVDVLSICVITFAMSFVPASVFLIOER 1680
 QY 1681 VSKAKHLOFISGVKPVLYLWLNFWDMCNYVVPATLVIIIFIOFOOKSYVSSNLPVIAL 1740
 DB 1681 VSKAKHLOFISGVKPVLYLWLNFWDMCNYVVPATLVIIIFIOFOOKSYVSSNLPVIAL 1740
 QY 1741 LLLYGMSITPFLMYPASFVEKIESTAVVLITSVNLFTGINGSVAFFLELFTDNKLNNTIN 1800

```

|||||
Db 1741 LLLXGWSITPLMYPASAEVFKIPSTAYVLTISVNLFIGINSVATFVLELFTDNKLNIN 1800
Qy 1801 DLKSVFLIFPHFCGRLIDVKNQANADALEREENRFPVSPISMDLVGRNLFMAVEG 1860
Db 1801 DLKSVFLIFPHFCGRLIDVKNQANADALEREENRFPVSPISMDLVGRNLFMAVEG 1860
Qy 1861 VVEFLITVLIQYRFRIRPPVNAKLSPLNDEDEDVREKORILDEGGONDIIEIKELTKI 1920
Db 1861 VVEFLITVLIQYRFRIRPPVNAKLSPLNDEDEDVREKORILDEGGONDIIEIKELTKI 1920
Qy 1921 YRRKRPAVDRIQYRFRIRPPVNAKLSPLNDEDEDVREKORILDEGGONDIIEIKELTKI 1980
Db 1921 YRRKRPAVDRIQYRFRIRPPVNAKLSPLNDEDEDVREKORILDEGGONDIIEIKELTKI 1980
Qy 1981 IHEVHONMGYCFQFADITELLTGREHVEFFALLRGVPEKEVGKGMATRKGLVKYGEK 2040
Db 1981 IHEVHONMGYCFQFADITELLTGREHVEFFALLRGVPEKEVGKGMATRKGLVKYGEK 2040
Qy 2041 YAGNYSGNKRKLTAMALIGGPVVFLEDEPTTGMDPKARFLMNCALSVEKGRSVLT 2100
Db 2041 YAGNYSGNKRKLTAMALIGGPVVFLEDEPTTGMDPKARFLMNCALSVEKGRSVLT 2100
Qy 2101 SHSMECEALCTRMATMNGRFRCLGSVOHLKNRFGDGTIYVRIAGSNPDLKPVODFTG 2160
Db 2101 SHSMECEALCTRMATMNGRFRCLGSVOHLKNRFGDGTIYVRIAGSNPDLKPVODFTG 2160
Qy 2161 LAPPGVSLKRRHNMLOYOLPSSLSSLARIFSLSSQSKRRLIEDYVSQOTLLDOYFVNF 2220
Db 2161 LAPPGVSLKRRHNMLOYOLPSSLSSLARIFSLSSQSKRRLIEDYVSQOTLLDOYFVNF 2220
Qy 2221 AKQSDDDHLKDLSLHKNTVVDVAVLTSFLQDEKYESYV 2261
Db 2221 AKQSDDDHLKDLSLHKNTVVDVAVLTSFLQDEKYESYV 2261

RESULT 2
AAB31365
ID AAB31365 standard; Protein: 2261 AA.
XX
AC AAB31365;
XX
DT 20-APR-2001 (first entry)
XX
DE Amino acid sequence of a human ABC1 polypeptide.
XX
KW Human: adenosine triphosphate binding cassette protein 1; ABC1;
KW apolipoprotein-mediated mobilisation; cholesterol; Tangier disease;
KW chromosome 9q22-9q31; heart disease; hypercholesterolemia;
KW atherosclerosis; cholesterol transport.
XX
OS Homo sapiens.
XX
PN WO200078971-A2.
XX
PD 28-DEC-2000.
XX
PE 16-JUN-2000; 2000WO-0516591.
XX
PR 18-JUN-1999; 99US-0140264.
PR 14-SEP-1999; 99US-0153872.
PR 19-NOV-1999; 99US-0166573.
XX
PA (CVTH-) CV THERAPEUTICS INC.
PA (UNIW ) UNIT WASHINGTON.
XX
PI Lawn RM, Wade D, Oram JF, Garvin M;
XX
DR WPI: 2001-137811/14.
DR N-PSDB: AAF24702.
XX
PT Adenosine triphosphate (ATP) binding cassette protein (ABC) 1
polynucleotides and polypeptides, useful for treatment of heart disease

```

```

PT and other disorders associated with hypercholesterolemia and
PT atherosclerosis -
XX
XX Claim 1: Page 123-138; 21pp; English.
XX
CC The present sequence represents a human adenosine triphosphate (ATP)
CC binding cassette protein (ABC) 1 polypeptide. ABC1 resides in cell
CC membranes and utilizes ATP hydrolysis to transport a wide variety of
CC substrates across the plasma membrane. ABC1 is a pivotal protein in
CC the apolipoprotein-mediated mobilisation of intracellular cholesterol
CC stores. ABC1 is defective in Tangier disease, a genetic disorder
CC characterised by abnormal HDL-cholesterol metabolism. The ABC1 gene is
CC localised to chromosome 9q22-9q31. The ABC1 genes and proteins are
CC useful for developing pharmaceutical agents for the treatment of heart
CC disease and other disorders associated with hypercholesterolemia and
CC atherosclerosis. The genes are useful for developing screening assays to
CC screen for compounds that regulate the expression of genes associated
CC with cholesterol transport. The genes and proteins are also useful for
CC are also useful as diagnostic indicators of cardiovascular disease and
CC other disorders associated with hypercholesterolemia.
XX
SQ Sequence 2261 AA:
Query Match 100.0%; Score 11797; DB 22; Length 2261;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MACWPQLRLMLKNTLFRROTCQLLEVPALFPLILISVRLSYPPEDQHECHFNKA 60
Db 1 MACWPQLRLMLKNTLFRROTCQLLEVPALFPLILISVRLSYPPEDQHECHFNKA 60
Qy 61 MPSAGTLRWVOCITICNANPCFRYPTPEAGCYVGNFNKSIVARLFSDARLLLYSQKDT 120
Db 61 MPSAGTLRWVOCITICNANPCFRYPTPEAGCYVGNFNKSIVARLFSDARLLLYSQKDT 120
Qy 121 SMKDMRKVLTLDQITKSSSNLKLQDFLVNDETFSGLFYHNLSPKSTVDMRLADYILH 180
Db 121 SMKDMRKVLTLDQITKSSSNLKLQDFLVNDETFSGLFYHNLSPKSTVDMRLADYILH 180
Qy 181 KYFLOGYQLHLTSLCNGSKSEEMTQLGDOEVSELGLPREKLAARERLRSNMDILKYL 240
Db 181 KYFLOGYQLHLTSLCNGSKSEEMTQLGDOEVSELGLPREKLAARERLRSNMDILKYL 240
Qy 241 RTLNSTSPPSKELAEATYTLTSLGTLAQLFMSRMSDMQEWMLFTNWNSSSSSTOI 300
Db 241 RTLNSTSPPSKELAEATYTLTSLGTLAQLFMSRMSDMQEWMLFTNWNSSSSSTOI 300
Qy 301 YQAVSRIVCGHPGGGLIKSLNMYEDNNYKALFGNGTEDEAFYDNSTTPYCNDLMK 360
Db 301 YQAVSRIVCGHPGGGLIKSLNMYEDNNYKALFGNGTEDEAFYDNSTTPYCNDLMK 360
Qy 361 NLESSPLSRIIMKALKPLLVGKILYTPDTPATROYVAEYNAKTFQELAVFHDEGWEELS 420
Db 361 NLESSPLSRIIMKALKPLLVGKILYTPDTPATROYVAEYNAKTFQELAVFHDEGWEELS 420
Qy 421 PKTWFMENSQENDLYRMLDSDRNDHFEQDLDGDMTAODIYVFLAKHPEDVOSSNGS 480
Db 421 PKTWFMENSQENDLYRMLDSDRNDHFEQDLDGDMTAODIYVFLAKHPEDVOSSNGS 480
Qy 481 VYTWREAFNETNOAIRTISFMECVNLKLEPATEVWLINSMELDERKFMAGIVFTG 540
Db 481 VYTWREAFNETNOAIRTISFMECVNLKLEPATEVWLINSMELDERKFMAGIVFTG 540
Qy 541 ITPGSIELPHVKKYKIRMDIDNVERTNKIKDGYWDFGPRADEDMRYWGGFAYLQDYV 600
Db 541 ITPGSIELPHVKKYKIRMDIDNVERTNKIKDGYWDFGPRADEDMRYWGGFAYLQDYV 600
Qy 601 EQALIRVLTGTEKKTGYVMOQMPYPCYVDIPLRWVSRSPMLPTLAMYISAVVITKGIY 660
Db 601 EQALIRVLTGTEKKTGYVMOQMPYPCYVDIPLRWVSRSPMLPTLAMYISAVVITKGIY 660
Qy 661 YEKARLKETMRIMGDINSILMFSWFTSSLPLVYAGLLVVLKGNLPLVSDPSVFEV 720
Db 661 YEKARLKETMRIMGDINSILMFSWFTSSLPLVYAGLLVVLKGNLPLVSDPSVFEV 720

```

Db 661 YEKEARLKEETMRJMGJLNDLSLWFSWFISSLLPDLVSAQLLVITLKGJLLEPSPSVFV 720
 QY 721 FLTSVFAVYVITLLOCELIISTLFSRANLAAAGGIIYFTLYLPVLCVAMODYVGTLLKIFAS 780
 Db 721 FLTSVFAVYVITLLOCELIISTLFSRANLAAAGGIIYFTLYLPVLCVAMODYVGTLLKIFAS 780
 QY 781 LLSVPAFGGCEFFALFEEOGIGVOMDLFESPVBEOSFNLTTSISMLLDTFLYGMVW 840
 Db 781 LLSVPAFGGCEFFALFEEOGIGVOMDLFESPVBEOSFNLTTSISMLLDTFLYGMVW 840
 QY 841 YLEAVFPGGYGIPRPWYPPCTKSYWFESEDESKSHPGSNOKRMEICMEEPHTLKGV 900
 Db 841 YLEAVFPGGYGIPRPWYPPCTKSYWFESEDESKSHPGSNOKRMEICMEEPHTLKGV 900
 QY 901 IONLVKVRDGMKAVDGLALNFEYGOTTSLFNGNAGKTTMSITLGLPPTSGTAYIL 960
 Db 901 IONLVKVRDGMKAVDGLALNFEYGOTTSLFNGNAGKTTMSITLGLPPTSGTAYIL 960
 QY 961 GKDISEMSTTRONLGYCPOHNVLFDMLTVEEHTMFYARLKGSEKHVKAEMOMALDVG 1020
 Db 961 GKDISEMSTTRONLGYCPOHNVLFDMLTVEEHTMFYARLKGSEKHVKAEMOMALDVG 1020
 QY 1021 LPSSKLKSTKSOLSGGMQRKLSVALAVYGSKVYITDEPLAGVDPYSRGTWELLKRYQ 1080
 Db 1021 LPSSKLKSTKSOLSGGMQRKLSVALAVYGSKVYITDEPLAGVDPYSRGTWELLKRYQ 1080
 QY 1081 GRTIILSTHMDADVLDGRIATISHGKLCGVSGSLFLKNOIGTYITLVKKEVSSLS 1140
 Db 1081 GRTIILSTHMDADVLDGRIATISHGKLCGVSGSLFLKNOIGTYITLVKKEVSSLS 1140
 QY 1141 SCRNSSTVSTLKKEDSVSSSDAGLGSDESHTLTIDVSAISNLTIRKHSFARLVEDI 1200
 Db 1141 SCRNSSTVSTLKKEDSVSSSDAGLGSDESHTLTIDVSAISNLTIRKHSFARLVEDI 1200
 QY 1201 GHELTLYVLPYFAKEGAFVELFHEIDRLSDLGISSETTLELFLKVAEESGVDAE 1260
 Db 1201 GHELTLYVLPYFAKEGAFVELFHEIDRLSDLGISSETTLELFLKVAEESGVDAE 1260
 QY 1261 TSDGTLPARRRRAFGDKQSCLRPFTEDDAADPNDSDIPESRETDLLSGMGKGSYQVK 1320
 Db 1261 TSDGTLPARRRRAFGDKQSCLRPFTEDDAADPNDSDIPESRETDLLSGMGKGSYQVK 1320
 QY 1321 GWRKLTQOOFVALLMKRLIARRSRKGFPAOIVPANEVCIALVFSIIVPFSKYPSELO 1380
 Db 1321 GWRKLTQOOFVALLMKRLIARRSRKGFPAOIVPANEVCIALVFSIIVPFSKYPSELO 1380
 QY 1381 PMWINEQYTFVSDAPEDTGTLELNALTRDPGFGTRCEGPNPIPTPCQAGEEEMTAP 1440
 Db 1381 PMWINEQYTFVSDAPEDTGTLELNALTRDPGFGTRCEGPNPIPTPCQAGEEEMTAP 1440
 QY 1441 VPOTIMDLFONGWMTQONPSPAQCQSSDKIKKMLPVCPPGAGGLPPPOKONTADILQDL 1500
 Db 1441 VPOTIMDLFONGWMTQONPSPAQCQSSDKIKKMLPVCPPGAGGLPPPOKONTADILQDL 1500
 QY 1501 TGRNISDYLVKTYVOTIAKSLKNKIWNEFEYGGFSLGVSNQALPPOEVDALIKOMK 1560
 Db 1501 TGRNISDYLVKTYVOTIAKSLKNKIWNEFEYGGFSLGVSNQALPPOEVDALIKOMK 1560
 QY 1561 HUKLADSSADRFNLNGLRPMGLDTRNNVYKWFNNKGWHAISFLNAINNAILRANLQK 1620
 Db 1561 HUKLADSSADRFNLNGLRPMGLDTRNNVYKWFNNKGWHAISFLNAINNAILRANLQK 1620
 QY 1621 GENPSHYGITAENHPLNTKQOLSEVALMTTSVDVLSICVIFAMSFPASVFFLQER 1680
 Db 1621 GENPSHYGITAENHPLNTKQOLSEVALMTTSVDVLSICVIFAMSFPASVFFLQER 1680
 QY 1681 VSKAKHJQFISGVKPYIWLNSFVDMCNVYVPATLVIIIFCPOKSYASTNLPVAL 1740
 Db 1681 VSKAKHJQFISGVKPYIWLNSFVDMCNVYVPATLVIIIFCPOKSYASTNLPVAL 1740
 QY 1741 LLLLYGKSTTIPLMYRASVYFKIPSTAYVVLVSVNLFIGINGSVATFVLELFTDKLNIN 1800
 Db 1741 LLLLYGKSTTIPLMYRASVYFKIPSTAYVVLVSVNLFIGINGSVATFVLELFTDKLNIN 1800

QY 1801 DLKSVPLIPPFHFCGRGLIDMVKNOAMADALEFFGENRVSPLSMDVGRNLFAMAABG 1860
 Db 1801 DLKSVPLIPPFHFCGRGLIDMVKNOAMADALEFFGENRVSPLSMDVGRNLFAMAABG 1860
 QY 1861 VFFELITVLIOYREFIRPRPVNAKLSPUNDEDEVRERORITIDGGQNDILEIKELTKI 1920
 Db 1861 VFFELITVLIOYREFIRPRPVNAKLSPUNDEDEVRERORITIDGGQNDILEIKELTKI 1920
 QY 1921 YRRKRKPADVDRICVIRPGECEGLGYNGAGKSSTEKMLTGTDTTVTRGDAFLNKNISLN 1980
 Db 1921 YRRKRKPADVDRICVIRPGECEGLGYNGAGKSSTEKMLTGTDTTVTRGDAFLNKNISLN 1980
 QY 1981 IHEVQNMNGYCPQPDATITELTGREHEFFALLRGVPEKVCVGMARLKIYKYGSK 2040
 Db 1981 IHEVQNMNGYCPQPDATITELTGREHEFFALLRGVPEKVCVGMARLKIYKYGSK 2040
 QY 2041 YAGNYSGGNKKKSTAMALIGGPPVFLDEPTTGMPKARFLMNCALSVEKGRSVLT 2100
 Db 2041 YAGNYSGGNKKKSTAMALIGGPPVFLDEPTTGMPKARFLMNCALSVEKGRSVLT 2100
 QY 2101 SHSMECEALCTRMAIWNNGRFCLGSVOHLKNRFGDGTIVYRTAGSNPDLKPYQDFEG 2160
 Db 2101 SHSMECEALCTRMAIWNNGRFCLGSVOHLKNRFGDGTIVYRTAGSNPDLKPYQDFEG 2160
 QY 2161 LAFPGSVLKEKRRNMLQYOLPSSLSIARIFSILSOSKRRLHIEDYSVQTTLDQVYNF 2220
 Db 2161 LAFPGSVLKEKRRNMLQYOLPSSLSIARIFSILSOSKRRLHIEDYSVQTTLDQVYNF 2220
 QY 2221 AKDOSDDHLKDLISLHKNOTVNVAVLTSTFLODEKESYV 2261
 Db 2221 AKDOSDDHLKDLISLHKNOTVNVAVLTSTFLODEKESYV 2261
 RESULT 3
 AAB38117
 ID AAB38117 standard; protein; 2261 AA.
 XX
 AC AAB38117;
 XX
 DT 29-JAN-2001 (first entry)
 XX
 DE Human ABC1 cholesterol transporter mutant, 1883M.
 XX
 KW Human ABC1 cholesterol transporter; chromosome 9q31;
 KW ATP-binding cassette; HDL deficiency disorder; high density lipoprotein;
 KW Tangier disease; TD; familial HDL deficiency; FNA; polymorphism;
 KW cardiovascular disease; coronary artery disease; coronary restenosis;
 KW cerebrovascular disease; peripheral vascular disease;
 KW Alzheimer's disease; Niemann-Pick disease; Huntington's disease;
 KW X-linked adrenoleukodystrophy; cancer; gene therapy; genetic diagnosis;
 KW prognosis; prophylaxis; drug screening; transgenic animal; mutant;
 KW
 OS Homo sapiens.
 XX
 PN WO20005318-A2.
 XX
 PD 21-SEP-2000.
 XX
 PF 15-MAR-2000; 2000WO-1B00532.
 XX
 PR 15-MAR-1999; 99US-0124702.
 PR 08-JUN-1999; 99US-0138048.
 PR 17-JUN-1999; 99US-0139600.
 PR 01-SEP-1999; 99US-0151977.
 XX
 PA (UYBR-) UNIV BRITISH COLUMBIA.
 PA (XENO-) XENON BIORESEARCH INC.
 XX
 PI Hayden MR, Wilson AR, Pimstone SN;
 XX WPI; 2000-587528/55.

XX New ABC1 polypeptide is useful for treating diseases associated with
 PT ABC1 biological activity, e.g. Alzheimer's disease, Huntington's
 PT disease and cancer -

PS Examples; Page - : 229pp; English.

XX The invention relates to the human ABC1 cholesterol transporter protein
 CC (B38082) and to nucleic acid sequences (C69120) which encode it. ABC1 is
 CC a member of the ATP-binding cassette (ABC transporter) superfamily of
 CC proteins, and plays a crucial role in cholesterol transport, particularly
 CC intracellular cholesterol trafficking in monocytes and fibroblasts, being
 CC involved in cholesterol efflux from the cell. The gene encoding ABC1 is
 CC located on chromosome 9q31, and mutations in this gene are associated
 CC with two genetic HDL (high density lipoprotein) deficiency disorders,
 CC Tangier disease (TD) and familial HDL deficiency (FHD). These diseases
 CC are distinguishable in that TD is an autosomal recessive disorder, while
 CC FHD is inherited as an autosomal dominant trait. Low levels of HDL ("good
 CC cholesterol") in the blood correlate with a high risk of cardiovascular
 CC disease, particularly coronary artery disease, but also cerebrovascular
 CC disease, coronary restenosis, and peripheral vascular disease.
 CC Conversely, a high level of HDL has protective effects against
 CC cardiovascular disease. The invention provides genetic constructs and
 CC transgenic cells and non-human animals comprising human ABC1 nucleic
 CC acids, and methods of gene therapy for the treatment or prevention of
 CC cardiovascular disease comprising the administration of an expression
 CC vector encoding ABC1 or an active fragment thereof. The invention also
 CC encompasses compounds which mimic ABC1 activity, compounds which
 CC stimulate ABC1 expression and methods of screening for such compounds.
 CC It further relates to methods for determining whether a patient has an
 CC increased risk for cardiovascular disease due to polymorphisms in the
 CC ABC1 gene. Human ABC1 proteins and nucleotides can be used to treat
 CC or prevent cardiovascular disease, especially coronary artery disease,
 CC cerebrovascular disease, coronary restenosis or peripheral vascular
 CC disease. They may also be used in the treatment of diseases associated
 CC with ABC1 biological activity, such as Alzheimer's disease, Niemann-Pick
 CC disease, Huntington's disease, X-linked adrenoleukodystrophy and cancer.
 CC The invention specifically excludes proteins with the exact amino acid
 CC sequences of Genbank Accession No: CAH10005.1 and X75926.1. The
 CC acid with the exact sequence as Genbank Accession No: AJ012376.1. The
 CC present sequence represents a mutant human ABC1 cholesterol transporter
 CC associated with an altered cholesterol level and therefore an altered
 CC risk of cardiovascular disease.
 CC Note: The present sequence is not shown in the specification, but is
 CC derived from the native human ABC1 shown on pages 152-157.

CC Sequence 2261 AA:

Query Match 100.0%; Score 11793; DB 21; Length 2261;

Best Local Similarity 99.9%; Pred. No. 0;

Matches 2259; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 MACWPQLRLMKLITRRRQTCOLLLEVAMPILFILILISVSLSPRYEOHCHCPNKA 60
 DB 1 MACWPQLRLMKLITRRRQTCOLLLEVAMPILFILILISVSLSPRYEOHCHCPNKA 60
 QY 61 MPAGTLPWAGIICNANPCFRYPTEGEAPGVGNFNKSIIVARLISDARILLYSQKDT 120
 DB 61 MPAGTLPWAGIICNANPCFRYPTEGEAPGVGNFNKSIIVARLISDARILLYSQKDT 120
 QY 121 SMKDKRVLRTLOQIKRSSSNLKLDELVDNETFSGLVYNLSLPKSTVDKMLRADYILH 180
 DB 121 SMKDKRVLRTLOQIKRSSSNLKLDELVDNETFSGLVYNLSLPKSTVDKMLRADYILH 180
 QY 181 KYFLOGYOLHLITSLCNGSKSEEMTOLGDOEYSELGLPKREKLAAEVYLSNMDILKPTL 240
 DB 181 KYFLOGYOLHLITSLCNGSKSEEMTOLGDOEYSELGLPKREKLAAEVYLSNMDILKPTL 240
 QY 241 RLUNSTSPFSPKEIAEAKTLLHSLGTLAQELFSMRSMQDMRQEMVLTJNVNSSSSSTOT 300
 DB 241 RLUNSTSPFSPKEIAEAKTLLHSLGTLAQELFSMRSMQDMRQEMVLTJNVNSSSSSTOT 300
 QY 301 YQAVSRIVCGHPREGGGLIKSLNLTEDNNYKALFGNGCTEDDATFTFDNSTTPYCNDLMK 360

DB 301 YQAVSRIVCGHPREGGGLIKSLNLTEDNNYKALFGNGCTEDDATFTFDNSTTPYCNDLMK 360
 QY 361 NLESSPLSRIRIMKALKPLLYGKILYTPDPTRQVMAEVKTFQELAVFHDLEGMWEELS 420
 DB 361 NLESSPLSRIRIMKALKPLLYGKILYTPDPTRQVMAEVKTFQELAVFHDLEGMWEELS 420
 QY 421 PKIWTFMENSOBMLVRLNLDLRNDHFVHQDLGDMTADIVAFIAKHPEDVOSSNGS 480
 DB 421 PKIWTFMENSOBMLVRLNLDLRNDHFVHQDLGDMTADIVAFIAKHPEDVOSSNGS 480
 QY 481 VYTWAEAFNETMOAIRTSRPEMECNLKLPIATEVVALINKSMELLDERKFWAGIVFTG 540
 DB 481 VYTWAEAFNETMOAIRTSRPEMECNLKLPIATEVVALINKSMELLDERKFWAGIVFTG 540
 QY 541 TTPGSTELPHVHKYKIRMDIDNERTNKIKGYNDPPGRADPFEDMRYVWGFAYLDDVY 600
 DB 541 TTPGSTELPHVHKYKIRMDIDNERTNKIKGYNDPPGRADPFEDMRYVWGFAYLDDVY 600
 QY 601 EQAIIIVLTGTEKKTGYVMQMPYPCYVDIIFLRMSKSMPLFTLAMIYSVAVILKIV 660
 DB 601 EQAIIIVLTGTEKKTGYVMQMPYPCYVDIIFLRMSKSMPLFTLAMIYSVAVILKIV 660
 QY 661 YEKEARLKEFMRIKGLDINSIMFSWFISSLPLPLVSAGILVILKGNLPLSDPSVVF 720
 DB 661 YEKEARLKEFMRIKGLDINSIMFSWFISSLPLPLVSAGILVILKGNLPLSDPSVVF 720
 QY 721 FLVSFAVNTILLOCFILSTFSRANLAACGGIYFTLPLVLCVAMODYGFILKIFAS 780
 DB 721 FLVSFAVNTILLOCFILSTFSRANLAACGGIYFTLPLVLCVAMODYGFILKIFAS 780
 QY 781 LLSPAVAFGCEYFALFEBOGIGVOMDLNLFESVPEDEFNLTSSIMLFDITLYGWTW 840
 DB 781 LLSPAVAFGCEYFALFEBOGIGVOMDLNLFESVPEDEFNLTSSIMLFDITLYGWTW 840
 QY 841 YIEAVPPOYGIIPRYWYFCOTSYWFGESDSKSPHSNOKMSEICMEFPYHLKIGVS 900
 DB 841 YIEAVPPOYGIIPRYWYFCOTSYWFGESDSKSPHSNOKMSEICMEFPYHLKIGVS 900
 QY 901 IONLVKVPDGMVAADGIALNFEEOGIRFGLHNGACKTTMTSILTGLPPTSGTAYIL 960
 DB 901 IONLVKVPDGMVAADGIALNFEEOGIRFGLHNGACKTTMTSILTGLPPTSGTAYIL 960
 QY 961 GKDIRSEMSTIRONLGYCPOHNVLFMLVBEIIVYALKGLSKHNAKAEQNALDVG 1020
 DB 961 GKDIRSEMSTIRONLGYCPOHNVLFMLVBEIIVYALKGLSKHNAKAEQNALDVG 1020
 QY 1021 LPSSKLSKTSOLSGMORLVSALAFVGSKYVILDEPTAGVDPSRSGIWEILLKTYRQ 1080
 DB 1021 LPSSKLSKTSOLSGMORLVSALAFVGSKYVILDEPTAGVDPSRSGIWEILLKTYRQ 1080
 QY 1081 GRTIILSTHHMDEADVLDGRIAIISHGKLCVGSSTFLKNOLGTGYLLTVKKDVESLS 1140
 DB 1081 GRTIILSTHHMDEADVLDGRIAIISHGKLCVGSSTFLKNOLGTGYLLTVKKDVESLS 1140
 QY 1141 SCRNSSTVSTYLRKEDSVSSDAGISDHESDPTLIDVSAISNLRKHVSEARLVEDI 1200
 DB 1141 SCRNSSTVSTYLRKEDSVSSDAGISDHESDPTLIDVSAISNLRKHVSEARLVEDI 1200
 QY 1201 GHELTLYVLPYAAKEGAFVELFHEIDRLSDLGISSYGISSETLLEIPLKVAEESGVDAE 1260
 DB 1201 GHELTLYVLPYAAKEGAFVELFHEIDRLSDLGISSYGISSETLLEIPLKVAEESGVDAE 1260
 QY 1261 TSDGTLPARRRNRAFGDKOSCLRPFTEDDAADPDSIDDESSETDLSGMDKSGSYOK 1320
 DB 1261 TSDGTLPARRRNRAFGDKOSCLRPFTEDDAADPDSIDDESSETDLSGMDKSGSYOK 1320
 QY 1321 GSKLQOQOFVALLMKRLLIARSRKGFPAQIVLPAVVCIALVPSLVPFGKYPSLELD 1380
 DB 1321 GSKLQOQOFVALLMKRLLIARSRKGFPAQIVLPAVVCIALVPSLVPFGKYPSLELD 1380
 QY 1381 PMVNYEQYTFVSNDAPEDTGTLLELNLTKDPCGTCMCEGNEPPTPCQAGEEEMVTAP 1440
 DB 1381 PMVNYEQYTFVSNDAPEDTGTLLELNLTKDPCGTCMCEGNEPPTPCQAGEEEMVTAP 1440

```

Db 1381 PMYNQEVTFVSNDAPEDTGTELELNLTKDPGFGTRCMENPIPTDPCQAGEEWTAP 1440
QY 1441 VFQITMDLFQNGNMTMOMPSPACQSSKIKKMLPVCPGAGGLPPQRRQNTADILQDL 1500
Db 1441 VFQITMDLFQNGNMTMOMPSPACQSSKIKKMLPVCPGAGGLPPQRRQNTADILQDL 1500
QY 1501 TGRNITSDYLKVTYVOIIKSKSLKNIMVNEFRYGGFSLGVSNTQALPSPQEVNATIKOMK 1560
Db 1501 TGRNITSDYLKVTYVOIIKSKSLKNIMVNEFRYGGFSLGVSNTQALPSPQEVNATIKOMK 1560
QY 1561 HIKLAKDSSADRFNLNSLGRFMTGLDTRNNVKVMEFNKGMHAISSEFLVINNAILRANLQK 1620
Db 1561 HIKLAKDSSADRFNLNSLGRFMTGLDTRNNVKVMEFNKGMHAISSEFLVINNAILRANLQK 1620
QY 1621 GENPSHYGTTAFNHPNLNTTKOQISEVALMTTSVDVLSICVITFAMSVVPSFVFLQER 1680
Db 1621 GENPSHYGTTAFNHPNLNTTKOQISEVALMTTSVDVLSICVITFAMSVVPSFVFLQER 1680
QY 1681 VSKAKHLOFISGVKPVYIWLNSFWDMCNVVPATLVIIIFICFOOKSVSSNTLPVAL 1740
Db 1681 VSKAKHLOFISGVKPVYIWLNSFWDMCNVVPATLVIIIFICFOOKSVSSNTLPVAL 1740
QY 1741 LLLLYGMSITPPLAMPASFVEKIPSTAVVLTSVNLFGLNGSVATFVLELFTDNKLNNTN 1800
Db 1741 LLLLYGMSITPPLAMPASFVEKIPSTAVVLTSVNLFGLNGSVATFVLELFTDNKLNNTN 1800
QY 1801 DILKSVLELPHFCLGRLIDMVKNOAMADALRFGENRFPVSLSDMLGRNLFMAVAG 1860
Db 1801 DILKSVLELPHFCLGRLIDMVKNOAMADALRFGENRFPVSLSDMLGRNLFMAVAG 1860
QY 1861 VVFLITVLIOYRPFIRPRVNAKLSPLNDEDDVRRQRILIDGGQNDILEIKELTKI 1920
Db 1861 VVFLITVLIOYRPFIRPRVNAKLSPLNDEDDVRRQRILIDGGQNDILEIKELTKI 1920
QY 1921 YRRRRKPAVDRIKCVGLPPECFLGLGVNAGKSTFKMLTGDITVTRGPAFLNKKSIISN 1980
Db 1921 YRRRRKPAVDRIKCVGLPPECFLGLGVNAGKSTFKMLTGDITVTRGPAFLNKKSIISN 1980
QY 1981 IHEHOMKGCYCPQDAITELTGREHVEFFALRGVPEKEVGKSEMAIRKLGIVKYEK 2040
Db 1981 IHEHOMKGCYCPQDAITELTGREHVEFFALRGVPEKEVGKSEMAIRKLGIVKYEK 2040
QY 2041 YAGNYSOGNKRKLTAMALIGPPVYLEDPTTGMDPKRRRLMCAISVKEGSSVLT 2100
Db 2041 YAGNYSOGNKRKLTAMALIGPPVYLEDPTTGMDPKRRRLMCAISVKEGSSVLT 2100
QY 2101 SHSMECEBALCTRMALMVNGRRCUGSVQHLKRRGDTYIVRIAGSNPDLKPVQDFG 2160
Db 2101 SHSMECEBALCTRMALMVNGRRCUGSVQHLKRRGDTYIVRIAGSNPDLKPVQDFG 2160
QY 2161 LAFPGSVLKEKHNNMLQYOLPSSLSIAATFSILSOSKRRLHIEDYSVSQTTLDQVNF 2220
Db 2161 LAFPGSVLKEKHNNMLQYOLPSSLSIAATFSILSOSKRRLHIEDYSVSQTTLDQVNF 2220
QY 2221 AKDQSDDDLKDLSLHKNQTVDAVLTSFLQDEKVEKESYV 2261
Db 2221 AKDQSDDDLKDLSLHKNQTVDAVLTSFLQDEKVEKESYV 2261

```

RESULT 4
AAB31362
ID AAB31362 standard; Protein: 2261 AA.

AC AAB31362;
DT 20-APR-2001 (first entry)

DE Amino acid sequence of ABC1 polypeptide from Tangier disease patient.

KW Human: adenosine triphosphate binding cassette protein 1; ABC1;

KW apolipoprotein-mediated mobilisation; cholesterol; Tangier disease;

KW chromosome 9q22-q931; heart disease; hypercholesterolemia;

KW atherosclerosis; cholesterol transport.

```

XX Homo sapiens.
OS
XX
XX Key Location/Qualifiers
FH Misc-difference 597
FT /note= "this is changed from Glu to Arg in Tangier
FT disease"
XX
XX W0200078972-A2.
XX
XX 28-DEC-2000.
XX
XX 16-JUN-2000; 2000WO-US16765.
XX
XX 18-JUN-1999; 99US-0140264.
XX 14-SEP-1999; 99US-0153872.
XX 19-NOV-1999; 99US-0166573.
XX
XX (CVTH-) CV THERAPEUTICS INC.
XX
XX Lawn RM, Wade D, Garvin M;
XX
XX WPI; 2001-137812/14.
XX
XX Adenosine triphosphate (ATP) binding cassette (ABC) polynucleotide,
XX useful for the development of agents for the treatment of heart disease
XX and other disorders associated with hypercholesterolemia and
XX atherosclerosis -
XX
XX Example 5; Page 154-169; 215pp; English.
XX
XX The present sequence represents a human adenosine triphosphate (ATP)
XX binding cassette protein (ABC) 1 polypeptide, and is isolated from
XX a Tangier disease patient. ABC1 resides in cell membranes and utilizes
XX ATP hydrolysis to transport a wide variety of substrates across the
XX plasma membrane. ABC1 is a pivotal protein in the apolipoprotein-mediated
XX mobilisation of intracellular cholesterol stores. ABC1 is defective in
XX Tangier disease, a genetic disorder characterised by abnormal
XX HDL-cholesterol metabolism. The ABC1 gene is localised to chromosome
XX 9q22-q931. The ABC1 genes and proteins are useful for developing
XX pharmaceutical agents for the treatment of heart disease and other
XX disorders associated with hypercholesterolemia and atherosclerosis. The
XX genes are useful for developing screening assays to screen for compounds
XX that regulate the expression of genes associated with cholesterol
XX transport. The genes and proteins are also useful for are also useful
XX as diagnostic indicators of cardiovascular disease and other disorders
XX associated with hypercholesterolemia.
XX
XX Sequence 2261 AA:
XX
XX Query Match 100.0%; Score 11793; DB 22; Length 2261;
XX Best Local Similarity 100.0%; Pred. No. 0;
XX Matches 2260; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1 MACWPQLRLMLKMLTFRROTCQLLEVAWPLFIPLILSVRLSYPPYDCHCHPNKA 60
XX 1 MACWPQLRLMLKMLTFRROTCQLLEVAWPLFIPLILSVRLSYPPYDCHCHPNKA 60
XX
XX 61 MPSACTLPWOGIICNANPCFRYPTEGAPGVGNFNKSIVARLESARLLYSQKDT 120
XX 61 MPSACTLPWOGIICNANPCFRYPTEGAPGVGNFNKSIVARLESARLLYSQKDT 120
XX
XX 121 SMKDMRKVLTLOQIKKSSNLKQDFLVNENPSSGLYHNLSLPKSTYDKMLRAVYLH 180
XX 121 SMKDMRKVLTLOQIKKSSNLKQDFLVNENPSSGLYHNLSLPKSTYDKMLRAVYLH 180
XX
XX 181 KYFLQGYOLHILTSICNGSKSEEMIQIGDOEVSSELGILPKKELAAARVLRSMNDILKPTL 240
XX 181 KYFLQGYOLHILTSICNGSKSEEMIQIGDOEVSSELGILPKKELAAARVLRSMNDILKPTL 240
XX
XX 241 RTLNSTSPFSPKELAEATKTLNLSLGLTLAGDELFSMRSWSDMRQEVAFITNVNSSSSSTQI 300
XX 241 RTLNSTSPFSPKELAEATKTLNLSLGLTLAGDELFSMRSWSDMRQEVAFITNVNSSSSSTQI 300

```

QY 301 YQAVSRIYCGHPEGGLIKISLNMVEDNNYKALRGNGTDEEDAEFTYDNSTTPYCNIDLMK 360
 Db 301 YQAVSRIYCGHPEGGLIKISLNMVEDNNYKALRGNGTDEEDAEFTYDNSTTPYCNIDLMK 360
 QY 361 NLESSPLRIITKALKPLLVGKILYPTPTPATROYMAEVNKTFOELAVFHLEGMWELS 420
 Db 361 NLESSPLRIITKALKPLLVGKILYPTPTPATROYMAEVNKTFOELAVFHLEGMWELS 420
 QY 421 PKIWTFMENSOEMDLVRLMLDSRDNDHFEQOOLDJMTAODIYAFKAKHEEDVOSSNGS 480
 Db 421 PKIWTFMENSOEMDLVRLMLDSRDNDHFEQOOLDJMTAODIYAFKAKHEEDVOSSNGS 480
 QY 481 VYTWDEARFETQOARTISREMECVNLKLEP IATEVWLKMSMLLDERKFMAGIVTGT 540
 Db 481 VYTWDEARFETQOARTISREMECVNLKLEP IATEVWLKMSMLLDERKFMAGIVTGT 540
 QY 541 TTPGSEILPHNHYKIKRMDIDNVERTNKIKDGYMDPCPRADPEEDMRYVMGCFAYLQDV 600
 Db 541 TTPGSEILPHNHYKIKRMDIDNVERTNKIKDGYMDPCPRADPEEDMRYVMGCFAYLQDV 600
 QY 601 EQAIIIRVLTEKTKGVYMQMPYCYVDJFLRVMSRSMPLMTLAMIYSAVVIKGIY 660
 Db 601 EQAIIIRVLTEKTKGVYMQMPYCYVDJFLRVMSRSMPLMTLAMIYSAVVIKGIY 660
 QY 661 YKEKARKETMRIMGLDNLMEFMSFISLIPLYSAGLIVYLKLGMLLPDSVYFV 720
 Db 661 YKEKARKETMRIMGLDNLMEFMSFISLIPLYSAGLIVYLKLGMLLPDSVYFV 720
 QY 721 FLTSFVAVVITLQCFILISTFSRANLAACGGLIYFTLYPYLYCYAAMDYVGTFTKIFAS 780
 Db 721 FLTSFVAVVITLQCFILISTFSRANLAACGGLIYFTLYPYLYCYAAMDYVGTFTKIFAS 780
 QY 781 LLSVPAFGCEYFALFEEQIGVQMDNLFESPBEDGFNLTTSSMILFDTFLYGMWT 840
 Db 781 LLSVPAFGCEYFALFEEQIGVQMDNLFESPBEDGFNLTTSSMILFDTFLYGMWT 840
 QY 841 YTEAVPFGQYGIIPRYFPCCTKSYMGEESDEKSHPGSNOKRMSFCMEBEPTHLKLGVS 900
 Db 841 YTEAVPFGQYGIIPRYFPCCTKSYMGEESDEKSHPGSNOKRMSFCMEBEPTHLKLGVS 900
 QY 901 IONLVKVRDGMKVAVDLALNFEQOITSPFLHGAGKTTMSILTGLPPTSGTAYIL 960
 Db 901 IONLVKVRDGMKVAVDLALNFEQOITSPFLHGAGKTTMSILTGLPPTSGTAYIL 960
 QY 961 GKDIREMSTIRONLGVCGQHNVLFMLTVEBEHIMFYARKGLSEKHVAAEMQALDVG 1020
 Db 961 GKDIREMSTIRONLGVCGQHNVLFMLTVEBEHIMFYARKGLSEKHVAAEMQALDVG 1020
 QY 1021 LPSKSLKSTLSQSGMOKRLSVALAFVGSKVYLDEPTAGVDPYSRGIMELLKTYRQ 1080
 Db 1021 LPSKSLKSTLSQSGMOKRLSVALAFVGSKVYLDEPTAGVDPYSRGIMELLKTYRQ 1080
 QY 1081 GRTIILSTHHMDEADVLDGRIALISHGLCCVGSSEFLKNQLGSTGYTLTVKKNDESSLS 1140
 Db 1081 GRTIILSTHHMDEADVLDGRIALISHGLCCVGSSEFLKNQLGSTGYTLTVKKNDESSLS 1140
 QY 1141 SCNNSSTVYLKKEEDSVQSSDAGLSDHESDTLTIVSAISNLIKRVHSAFARLVEDI 1200
 Db 1141 SCNNSSTVYLKKEEDSVQSSDAGLSDHESDTLTIVSAISNLIKRVHSAFARLVEDI 1200
 QY 1201 GHEILYVLPYEAKEGAFVELFHEIDRLSDLGISSYGISETTLEEIFKVAEESGVDAE 1260
 Db 1201 GHEILYVLPYEAKEGAFVELFHEIDRLSDLGISSYGISETTLEEIFKVAEESGVDAE 1260
 QY 1261 TSDGTLPARNRARAFGDKOSCLRPFTEDDAADPNDSIDPESEFTDLVSGMCKGKYQV 1320
 Db 1261 TSDGTLPARNRARAFGDKOSCLRPFTEDDAADPNDSIDPESEFTDLVSGMCKGKYQV 1320
 QY 1321 GMLTQOQFVALLMKRLILARRSRKGFPAQIYLPAYFVCIALVFSLIVPEFGKTPSLEIO 1380
 Db 1321 GMLTQOQFVALLMKRLILARRSRKGFPAQIYLPAYFVCIALVFSLIVPEFGKTPSLEIO 1380

QY 1381 PMWNEQYFVNSDADPEDTLELLNALITKDPGFGRCMEGNPIDPTPCOAGEEEMTAP 1440
 Db 1381 PMWNEQYFVNSDADPEDTLELLNALITKDPGFGRCMEGNPIDPTPCOAGEEEMTAP 1440
 QY 1441 VPQTIMDLFONGMWTMONPSPACQCSSDKIKKMLPVCPPGAGLPPPOKKONTADILQDL 1500
 Db 1441 VPQTIMDLFONGMWTMONPSPACQCSSDKIKKMLPVCPPGAGLPPPOKKONTADILQDL 1500
 QY 1501 TGRNISDYLVKTYVOITAKSLKKNIVNEFRYGGFSLGVSNTQALPPSOEVNDAIKOMK 1560
 Db 1501 TGRNISDYLVKTYVOITAKSLKKNIVNEFRYGGFSLGVSNTQALPPSOEVNDAIKOMK 1560
 QY 1561 HLKAKDSSADRELNSGRMTGLDTRNNKVMFNKKGMHAISSPLNVLINNALIRANLQK 1620
 Db 1561 HLKAKDSSADRELNSGRMTGLDTRNNKVMFNKKGMHAISSPLNVLINNALIRANLQK 1620
 QY 1621 GENSHYGIATFANHPLNLTKQOLSEVALMTTSYDVLSICVIFAMSFPASFPVFLIOER 1680
 Db 1621 GENSHYGIATFANHPLNLTKQOLSEVALMTTSYDVLSICVIFAMSFPASFPVFLIOER 1680
 QY 1681 VSKAKHLOFTSGVPVLYWLSNFWQMCNTVVPATLVIIIFICFOOKSYVSTNLPLVL 1740
 Db 1681 VSKAKHLOFTSGVPVLYWLSNFWQMCNTVVPATLVIIIFICFOOKSYVSTNLPLVL 1740
 QY 1741 LLLLYGMSITPPLMYPASFEVERIPSTAVVVLTSVNLFIGINGSVATEVLELTDNKLNIN 1800
 Db 1741 LLLLYGMSITPPLMYPASFEVERIPSTAVVVLTSVNLFIGINGSVATEVLELTDNKLNIN 1800
 QY 1801 DILKSVFLIPHECLRGCLIDMVKNQAMADALERFGENRFPVSPSLMDVGRNLFMAAABG 1860
 Db 1801 DILKSVFLIPHECLRGCLIDMVKNQAMADALERFGENRFPVSPSLMDVGRNLFMAAABG 1860
 QY 1861 VVEFLITVLYOYREFIRPRPVNAKLSPLNDEDEVRERORIIIDGGGQNDILEIKELTKI 1920
 Db 1861 VVEFLITVLYOYREFIRPRPVNAKLSPLNDEDEVRERORIIIDGGGQNDILEIKELTKI 1920
 QY 1921 YRRKRKPAVDRIQVIGIPGECFGLLGYNAGKSSFTFMTLGTDTVTTPGDAFLKNKSTLSN 1980
 Db 1921 YRRKRKPAVDRIQVIGIPGECFGLLGYNAGKSSFTFMTLGTDTVTTPGDAFLKNKSTLSN 1980
 QY 1981 IHEVHQNMGYCPQEDATITELLTGREHVEFPALLRGVBEKYGKGEWAIRKLGIVKXGK 2040
 Db 1981 IHEVHQNMGYCPQEDATITELLTGREHVEFPALLRGVBEKYGKGEWAIRKLGIVKXGK 2040
 QY 2041 YAGNYSQGNKRKRLSTAMALIGCPVVFLEDEPTGMDKARFPLNCALSVYKEGRSVLT 2100
 Db 2041 YAGNYSQGNKRKRLSTAMALIGCPVVFLEDEPTGMDKARFPLNCALSVYKEGRSVLT 2100
 QY 2101 SHSMECEALCTRNAIMVNGRFCLGSVOHLKNRFGDGYTTVRIAGSNPDLKPVODFG 2160
 Db 2101 SHSMECEALCTRNAIMVNGRFCLGSVOHLKNRFGDGYTTVRIAGSNPDLKPVODFG 2160
 QY 2161 LAFPSVYLKEKHNRLQVQLPSSLSLARIFSILISQSKRRIHEDYSVQTTLDQVVFNF 2220
 Db 2161 LAFPSVYLKEKHNRLQVQLPSSLSLARIFSILISQSKRRIHEDYSVQTTLDQVVFNF 2220
 QY 2221 AKDQSDDDLKDLSLHKNQTVVDAVLTSLFQDEKVESYV 2261
 Db 2221 AKDQSDDDLKDLSLHKNQTVVDAVLTSLFQDEKVESYV 2261

RESULT 5
 AAB31366
 ID AAB31366 standard; Protein: 2261 AA.
 XX
 AC AAB31366:
 XX
 DT 20-Apr-2001 (first entry)
 XX
 DE Amino acid sequence of ABC1 polypeptide from Tangier disease patient.
 XX Human: adenosine triphosphate binding cassette protein 1; ABC1;
 KM apolipoprotein-mediated mobilisation; cholesterol; Tangier disease;

KM chromosome 9q22-q31; heart disease; hypercholesterolemia;
 KM atherosclerosis; cholesterol transport.
 XX
 XX Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 597 /note="this is changed from Glu to Arg in Tangler
 FT disease"
 FT
 WO200078971-AZ.
 XX
 XX 28-DEC-2000.
 XX
 XX 16-JUN-2000; 2000WO-0S16591.
 XX
 XX 18-JUN-1999; 99US-0140264.
 PR 14-SEP-1999; 99US-0153872.
 PR 19-NOV-1999; 99US-0166573.
 XX
 PA (CUTH-) CV THERAPEUTICS INC.
 PA (UNITW) UNITV WASHINGTON.
 XX
 PI Lawn RM, Wade D, Oram JF, Garvin M;
 XX
 DR WPI: 2001-137811/14.
 DR N-PSDB: AAF24707.
 XX
 PT Adenosine triphosphate (ATP) binding cassette protein (ABC) 1
 PT polynucleotides and polypeptides, useful for treatment of heart disease
 PT and other disorders associated with hypercholesterolemia and
 PT atherosclerosis -
 XX
 PS Claim 25; Page 150-165; 211pp; English.
 XX
 CC The present sequence represents a human adenosine triphosphate (ATP)
 CC binding cassette protein (ABC) 1 polypeptide, and is isolated from
 CC a Tangler disease patient. ABC1 resides in cell membranes and utilizes
 CC ATP hydrolysis to transport a wide variety of substrates across the
 CC plasma membrane. ABC1 is a pivotal protein in the apolipoprotein-mediated
 CC mobilisation of intracellular cholesterol stores. ABC1 is defective in
 CC Tangler disease, a genetic disorder characterised by abnormal
 CC HDL-cholesterol metabolism. The ABC1 gene is localised to chromosome
 CC 9q22-q31. The ABC1 genes and proteins are useful for developing
 CC pharmaceutical agents for the treatment of heart disease and other
 CC disorders associated with hypercholesterolemia and atherosclerosis. The
 CC genes are useful for developing screening assays to screen for compounds
 CC that regulate the expression of genes associated with cholesterol
 CC transport. The genes and proteins are also useful for cholesterol
 CC as diagnostic indicators of cardiovascular disease and other disorders
 CC associated with hypercholesterolemia.
 CC
 XX Sequence 2261 AA:
 SQ
 Query Match 100.0%; Score 11793; DB 22; Length 2261;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2260; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 241 RLINSTSPFSKELAEATKTLTSLGTLAOLFLSNBSWSMDROEVLINNVSSSSSTOI 300
 |||||||
 Db 241 RLINSTSPFSKELAEATKTLTSLGTLAOLFLSNBSWSMDROEVLINNVSSSSSTOI 300
 QY 301 YQAVSRIVCGHPEGGGIKIKSLMWEDNNYKALFGNGTEDEAEFFYDNTSTPYCNDLMK 360
 |||||||
 Db 301 YQAVSRIVCGHPEGGGIKIKSLMWEDNNYKALFGNGTEDEAEFFYDNTSTPYCNDLMK 360
 QY 361 NLESSPLSRITIMKALPVLVGLITPTPDTPATROVAEYVNTKFOELAVHDELMGEELS 420
 |||||||
 Db 361 NLESSPLSRITIMKALPVLVGLITPTPDTPATROVAEYVNTKFOELAVHDELMGEELS 420
 QY 421 PKITFEMENSQEMDLVRMLDSDNDHFEEOQLDGLMDTADIVAFIAHPEDVOSSNGS 480
 |||||||
 Db 421 PKITFEMENSQEMDLVRMLDSDNDHFEEOQLDGLMDTADIVAFIAHPEDVOSSNGS 480
 QY 481 VYTWREAFNETNOAIRTISRMECVNLNKLPIATEVWLINSMELDERKFWAGITFTG 540
 |||||||
 Db 481 VYTWREAFNETNOAIRTISRMECVNLNKLPIATEVWLINSMELDERKFWAGITFTG 540
 QY 541 ITPGSEIEPHHYVYKTRMDIDNVERTNKLKDGWDGPRADPEEDMRVYWGGAFLVDVY 600
 |||||||
 Db 541 ITPGSEIEPHHYVYKTRMDIDNVERTNKLKDGWDGPRADPEEDMRVYWGGAFLVDVY 600
 QY 601 EQAIRVLTGTEKKTGVYMOQMPYPCYVDIFLRVMSRSMPLMTLAWIYSAVITIKIV 660
 |||||||
 Db 601 EQAIRVLTGTEKKTGVYMOQMPYPCYVDIFLRVMSRSMPLMTLAWIYSAVITIKIV 660
 QY 661 YEKEARKETMRIMGIDNSILMFSWFTSSILPLVSAGILVYLKGNLPLPSDPSVVFV 720
 |||||||
 Db 661 YEKEARKETMRIMGIDNSILMFSWFTSSILPLVSAGILVYLKGNLPLPSDPSVVFV 720
 QY 721 FLVSFAVVTILLOCFILSTFESRANLAACGGITVFLLYLYCYAAMDYVFTLKITPAS 780
 |||||||
 Db 721 FLVSFAVVTILLOCFILSTFESRANLAACGGITVFLLYLYCYAAMDYVFTLKITPAS 780
 QY 781 LLSVPAFVGCGCEYFALFEEOGIGVQMDNLFESPVEEDGNLTTSISMMLFDTLXGVTW 840
 |||||||
 Db 781 LLSVPAFVGCGCEYFALFEEOGIGVQMDNLFESPVEEDGNLTTSISMMLFDTLXGVTW 840
 QY 841 YIYAVPVGQYGIYPRPYFECTSYWFGFESDEKSHPGSNOKMSTICHEEPTHLKGLVS 900
 |||||||
 Db 841 YIYAVPVGQYGIYPRPYFECTSYWFGFESDEKSHPGSNOKMSTICHEEPTHLKGLVS 900
 QY 901 IONLVKVRBDGKAVAVDGLALNFYEGQITSPFHNGAGKTTMTSILTGLFPTSGTAYIL 960
 |||||||
 Db 901 IONLVKVRBDGKAVAVDGLALNFYEGQITSPFHNGAGKTTMTSILTGLFPTSGTAYIL 960
 QY 961 GMDIREKMTIQONLGVCPQHNVLFDMLTVEHIMFYARLKLGLSKHYKAEBQMALDVG 1020
 |||||||
 Db 961 GMDIREKMTIQONLGVCPQHNVLFDMLTVEHIMFYARLKLGLSKHYKAEBQMALDVG 1020
 QY 1021 LPSSKLSKSTQSLSGMOKRISVALAFVGSKVYLLDPTGAGVDPYSRGRIWELLKLRQ 1080
 |||||||
 Db 1021 LPSSKLSKSTQSLSGMOKRISVALAFVGSKVYLLDPTGAGVDPYSRGRIWELLKLRQ 1080
 QY 1081 GRTIILSTHMDADVLDGRIALISHGKLCCVSSILFLKNLGTGYVYLLVKKKIVESSIS 1140
 |||||||
 Db 1081 GRTIILSTHMDADVLDGRIALISHGKLCCVSSILFLKNLGTGYVYLLVKKKIVESSIS 1140
 QY 1141 SCRNSSTVSYLTKKEDSVSSQSSDAGLSDHESDPLTIDVSAISNLIRKHVSEARLVEDI 1200
 |||||||
 Db 1141 SCRNSSTVSYLTKKEDSVSSQSSDAGLSDHESDPLTIDVSAISNLIRKHVSEARLVEDI 1200
 QY 1201 GHELTIVYLYPEAKAGAVVELFHEITDRLSDGISYSYISSETTLLEELFLKVAEESGDAD 1260
 |||||||
 Db 1201 GHELTIVYLYPEAKAGAVVELFHEITDRLSDGISYSYISSETTLLEELFLKVAEESGDAD 1260
 QY 1261 TSDGTLPARNRNRAFQKOSCLRPFTEDDADPNDSDIDPESRETDLISGMDGSGSYOVK 1320
 |||||||
 Db 1261 TSDGTLPARNRNRAFQKOSCLRPFTEDDADPNDSDIDPESRETDLISGMDGSGSYOVK 1320


```

OY 1321 GWKLTQOQFVALLMKRLIARRSRKGFPAQIVLPAVFVIALVPSLIVPFGKYPSLEQ 1380
    |||||||
DB 1321 GWKLTQOQFVALLMKRLIARRSRKGFPAQIVLPAVFVIALVPSLIVPFGKYPSLEQ 1380
OY 1381 PMWNEQYTFVSNDAPEDTGTLELLNLTGPGFGRCEMGNIPDPFQAGEEEMTAP 1440
    |||||||
DB 1381 PMWNEQYTFVSNDAPEDTGTLELLNLTGPGFGRCEMGNIPDPFQAGEEEMTAP 1440
OY 1441 VPQITMDLFQNGNWTQNPSPACQSSDKIKMLPVCPGAGGLPPQKQKNTADIIQDL 1500
    |||||||
DB 1441 VPQITMDLFQNGNWTQNPSPACQSSDKIKMLPVCPGAGGLPPQKQKNTADIIQDL 1500
OY 1501 TGRNISDVLVYTYVOIILKSLKNKIWNEFRYGGFSLGVSNTQALPPEQVNDATIKQKK 1560
    |||||||
DB 1501 TGRNISDVLVYTYVOIILKSLKNKIWNEFRYGGFSLGVSNTQALPPEQVNDATIKQKK 1560
OY 1561 HLKLAQSSADREFLNSLGRFMTGDTNRNNVWVFNNKGMHAISSEFLVNNAILRANLQK 1620
    |||||||
DB 1561 HLKLAQSSADREFLNSLGRFMTGDTNRNNVWVFNNKGMHAISSEFLVNNAILRANLQK 1620
OY 1621 GENPSHYGITAFNHPNLNTKQOLSVALMTTSVDLVISICVIFAMSPFVAFVFLIOER 1680
    |||||||
DB 1621 GENPSHYGITAFNHPNLNTKQOLSVALMTTSVDLVISICVIFAMSPFVAFVFLIOER 1680
OY 1681 VSKAKHLOFISGVKPVYIWLSENFVMDMCNYYVPATLVIIIFCFQOKSYVSTNLPLYAL 1740
    |||||||
DB 1681 VSKAKHLOFISGVKPVYIWLSENFVMDMCNYYVPATLVIIIFCFQOKSYVSTNLPLYAL 1740
OY 1741 LLLLYGWSITPLMPASFAVFKIPSTAYVVLTSVNLFIQINGSVATFVLEFETDNKLNIN 1800
    |||||||
DB 1741 LLLLYGWSITPLMPASFAVFKIPSTAYVVLTSVNLFIQINGSVATFVLEFETDNKLNIN 1800
OY 1801 DLKSVFLIFPHFCGRLIDVKNQAMADALERGENRPVSPSLMDLVGRNLPAMAVEG 1860
    |||||||
DB 1801 DLKSVFLIFPHFCGRLIDVKNQAMADALERGENRPVSPSLMDLVGRNLPAMAVEG 1860
OY 1861 VVEFLITVLIOYRFFIRRPVNAKLSPLNDEDEDEVRERORILLDGGGNDILEIKELTKI 1920
    |||||||
DB 1861 VVEFLITVLIOYRFFIRRPVNAKLSPLNDEDEDEVRERORILLDGGGNDILEIKELTKI 1920
OY 1921 YRRKKRPVADRICVGIPECECGLLGVGAGKSSFFKMLTGTITYTRGDAFLNKNSILSN 1980
    |||||||
DB 1921 YRRKKRPVADRICVGIPECECGLLGVGAGKSSFFKMLTGTITYTRGDAFLNKNSILSN 1980
OY 1981 HEHVONMGYCPQPAITPFLTGRHEVEFPALLRGVPEKVGKVEMAIRKLGIVKYEK 2040
    |||||||
DB 1981 HEHVONMGYCPQPAITPFLTGRHEVEFPALLRGVPEKVGKVEMAIRKLGIVKYEK 2040
OY 2041 YAGNYSNGKRRKLTAMALIGPPVFLDEPTTGMDPKARFLMNCALSVYEGRSVYL 2100
    |||||||
DB 2041 YAGNYSNGKRRKLTAMALIGPPVFLDEPTTGMDPKARFLMNCALSVYEGRSVYL 2100
OY 2101 SHSMECEALCTRAIMVNGRRRCISVOHLKNRFGDGTIYVRLAGSNPDLKPVQDFEG 2160
    |||||||
DB 2101 SHSMECEALCTRAIMVNGRRRCISVOHLKNRFGDGTIYVRLAGSNPDLKPVQDFEG 2160
OY 2161 LAFPGSVLKEKRRNMLOVLPSSLSLARIFSLISQSKRLHIEIYSVQSTLDOVFVNF 2220
    |||||||
DB 2161 LAFPGSVLKEKRRNMLOVLPSSLSLARIFSLISQSKRLHIEIYSVQSTLDOVFVNF 2220
OY 2221 AKDOSDDHKLKDLKLNKQTVVDVAVLTSFLQDEKVEKESYV 2261
    |||||||
DB 2221 AKDOSDDHKLKDLKLNKQTVVDVAVLTSFLQDEKVEKESYV 2261

```

```

DE Human ABC1 cholesterol transporter mutant, R219K.
XX
KW Human ABC1 cholesterol transporter; chromosome 9q31;
KW ATP-binding cassette; HDL deficiency disorder; high density lipoprotein;
KW Tangier disease; TD; familial HDL deficiency; FHA; polymorphism;
KW cardiovascular disease; coronary artery disease; coronary restenosis;
KW cerebrovascular disease; peripheral vascular disease;
KW Alzheimer's disease; Niemann-Pick disease; Huntington's disease;
KW X-linked adrenoleukodystrophy; cancer; gene therapy; genetic diagnosis;
KW prognosis; prophylaxis; drug screening; transgenic animal; mutant;
KW mutin.
XX
OS Homo sapiens.
XX
PN W0200055318-A2.
XX
PD 21-SEP-2000.
XX
PE 15-MAR-2000; 2000MO-IB00532.
XX
PR 15-MAR-1999; 99US-0124702.
PR 08-JUN-1999; 99US-0138048.
PR 17-JUN-1999; 99US-0139600.
PR 01-SEP-1999; 99US-0151977.
XX
PA (UYBR-) UNIV BRITISH COLUMBIA.
PA (XENO-) XENON BIORESEARCH INC.
XX
PI Hayden MR, Wilson AR, Pimstone SN.
XX
DR WPL; 2000-587528/55.
XX
PT New ABC1 polypeptide is useful for treating diseases associated with
PT ABC1 biological activity, e.g. Alzheimer's disease, Huntington's
PT disease and cancer -
XX
PS Examples: Page -: 229pp; English.
XX
CC The invention relates to the human ABC1 cholesterol transporter protein
CC (B38082) and to nucleic acid sequences (C69120) which encode it. ABC1 is
CC a member of the ATP-binding cassette (ABC transporter) superfamily of
CC proteins, and plays a crucial role in cholesterol transport, particularly
CC intracellular cholesterol trafficking in monocytes and fibroblasts, being
CC involved in cholesterol efflux from the cell. The gene encoding ABC1 is
CC located on chromosome 9q31, and mutations in this gene are associated
CC with two genetic HDL (high density lipoprotein) deficiency disorders.
CC Tangier disease (TD) and familial HDL deficiency (FHA). These diseases
CC are distinguishable in that TD is an autosomal recessive disorder, while
CC FHA is inherited as an autosomal dominant trait. Low levels of HDL ("good
CC cholesterol") in the blood correlate with a high risk of cardiovascular
CC disease, particularly coronary artery disease, but also cerebrovascular
CC disease, coronary restenosis, and peripheral vascular disease.
CC Conversely, a high level of HDL has protective effects against
CC cardiovascular disease. The invention provides genetic constructs and
CC transgenic cells and non-human animals comprising human ABC1 nucleic
CC acids, and methods of gene therapy for the treatment or prevention of
CC cardiovascular disease comprising the administration of an expression
CC vector encoding ABC1 or an active fragment thereof. The invention also
CC encompasses compounds which mimic ABC1 activity, compounds which
CC stimulate ABC1 expression and methods of screening for such compounds.
CC It further relates to methods for determining whether a patient has an
CC increased risk for cardiovascular disease due to polymorphisms in the
CC ABC1 gene. Human ABC1 proteins and nucleotides can be used to treat
CC or prevent cardiovascular disease, especially coronary artery disease,
CC cerebrovascular disease, coronary restenosis or peripheral vascular
CC disease. They may also be used in the treatment of diseases associated
CC with ABC1 biological activity, such as Alzheimer's disease, Niemann-Pick
CC disease, Huntington's disease, X-linked adrenoleukodystrophy and cancer.
CC The invention specifically excludes proteins with the exact amino acid
CC sequences of GenBank Accession No: CAA10005.1 and X75926, and the nucleic
CC acid with the exact sequence as GenBank Accession No: AJ012376.1. The
CC present sequence represents a mutant human ABC1 cholesterol transporter
CC associated with an altered cholesterol level and therefore an altered

```

CC Risk of cardiovascular disease.

CC Note: The present sequence is not shown in the specification, but is derived from the native human ABC1 shown on pages 152-157.

XX Sequence 2261 AA:

Query Match 100.0%; Score 11792; DB 21; Length 2261;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2259; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 MACMPQLRLMLKNTLFRROTCTLLLEAVMPLEFLITLISVRLSYPEYEOHECFPKKA 60
Db 1 MACMPQLRLMLKNTLFRROTCTLLLEAVMPLEFLITLISVRLSYPEYEOHECFPKKA 60
QY 61 MSAGTLPLVVOGICTANNPCFRYPYPCGAPGVGNFKSIVARLFSDARILLYSQDRT 120
Db 61 MSAGTLPLVVOGICTANNPCFRYPYPCGAPGVGNFKSIVARLFSDARILLYSQDRT 120
QY 121 SKMDMKVRLTLOQIKKSSSNLKLODPLVDNETPSGLYHNLSPKSTVYDKMLRADVILH 180
Db 121 SKMDMKVRLTLOQIKKSSSNLKLODPLVDNETPSGLYHNLSPKSTVYDKMLRADVILH 180
QY 181 KVFLOGYQLHLTSLCNGSKSEEMIQLDQEVSELCGLPKKELAAAEVRLSRMMDILKPI 240
Db 181 KVFLOGYQLHLTSLCNGSKSEEMIQLDQEVSELCGLPKKELAAAEVRLSRMMDILKPI 240
QY 241 RTLNSTSPRSKELAEAKRTLLHSLGTLAQLFESMRSSMDKQEVMLTNNSSSSSTQI 300
Db 241 RTLNSTSPRSKELAEAKRTLLHSLGTLAQLFESMRSSMDKQEVMLTNNSSSSSTQI 300
QY 301 YQAVSRIVCGHEGGGLKISLWYEDNNYKALFGNGT EEDAEFYNNSTTPYCNDLML 360
Db 301 YQAVSRIVCGHEGGGLKISLWYEDNNYKALFGNGT EEDAEFYNNSTTPYCNDLML 360
QY 361 NLESSPLSRITIKALKPLLYGKILYPTDPTATROYMAEVNTPFOLAVFHDLLEGMEELS 420
Db 361 NLESSPLSRITIKALKPLLYGKILYPTDPTATROYMAEVNTPFOLAVFHDLLEGMEELS 420
QY 421 PKIWFEMESQEMDLVRLMLDSRDNDHFEQOOLGDMTADIVAFILAKHEDVOSSSGS 480
Db 421 PKIWFEMESQEMDLVRLMLDSRDNDHFEQOOLGDMTADIVAFILAKHEDVOSSSGS 480
QY 481 VYTWREAFNETNOALIRTSREMECVNLKLEPIATEVWLINKSMELDEREFMAGIVFTG 540
Db 481 VYTWREAFNETNOALIRTSREMECVNLKLEPIATEVWLINKSMELDEREFMAGIVFTG 540
QY 541 ITPGSTELPHHYKXKIRMDINDVERTNKIKGYMDPGPRADPFEDMRYVWGCFAYLQDV 600
Db 541 ITPGSTELPHHYKXKIRMDINDVERTNKIKGYMDPGPRADPFEDMRYVWGCFAYLQDV 600
QY 601 EQAIIIRVLGTEKKTGVYMOOMPYPYCYDDJFLRYMSRSMPLFMILAMTYSVAVILKIV 660
Db 601 EQAIIIRVLGTEKKTGVYMOOMPYPYCYDDJFLRYMSRSMPLFMILAMTYSVAVILKIV 660
QY 661 YEKARLKEKTMIRMGDLSILMFSSFTSLIPLLYSAGLLVILKGLNLPYSDPSVVEV 720
Db 661 YEKARLKEKTMIRMGDLSILMFSSFTSLIPLLYSAGLLVILKGLNLPYSDPSVVEV 720
QY 721 FLVSFAVVTIIIOCLIFLISLFSRANLAAAGGIIYFTLYLPRVLCVANOVDYGFLLKIFAS 780
Db 721 FLVSFAVVTIIIOCLIFLISLFSRANLAAAGGIIYFTLYLPRVLCVANOVDYGFLLKIFAS 780
QY 781 LLSPAVFGGCEYFALFEEOGIGVOMDNLFESEPEEDFNLITTSISMFLFTFLYGVMTW 840
Db 781 LLSPAVFGGCEYFALFEEOGIGVOMDNLFESEPEEDFNLITTSISMFLFTFLYGVMTW 840
QY 841 YIEAFVPGQYGIIPRPWTFCSTWFGESDEKSHPGSNOKRMSEICNEEPTHLKLGVS 900
Db 841 YIEAFVPGQYGIIPRPWTFCSTWFGESDEKSHPGSNOKRMSEICNEEPTHLKLGVS 900
QY 901 IONLVKVRDGMKVAVDGLALNFYEGQITSLFGHNGAGKTTMSILGLPPPTSGTAYIL 960
Db 901 IONLVKVRDGMKVAVDGLALNFYEGQITSLFGHNGAGKTTMSILGLPPPTSGTAYIL 960

QY 961 GKDIRSEMSIIRONLGYCPOHNVLFDMIVVEEHIMFARLKLSEKHVKAEMFOMALDVG 1020
Db 961 GKDIRSEMSIIRONLGYCPOHNVLFDMIVVEEHIMFARLKLSEKHVKAEMFOMALDVG 1020
QY 1021 LPSSKLKSKTSQSLSGMQRKLSVALAFVGSKVYLIDEPTAGVDPYSRNGIWEILLKRYRQ 1080
Db 1021 LPSSKLKSKTSQSLSGMQRKLSVALAFVGSKVYLIDEPTAGVDPYSRNGIWEILLKRYRQ 1080
QY 1081 GRTITLSTHMHMDEADVIGDGLIISHGKILCCVSSSLFIKNQIGTGYLTLTKKDVSSIS 1140
Db 1081 GRTITLSTHMHMDEADVIGDGLIISHGKILCCVSSSLFIKNQIGTGYLTLTKKDVSSIS 1140
QY 1141 SCRNSSTVSYLAKKEDSVSSQSSDAGLSGDSHEDTLTIDVSAISNLRKHVSARLVEDI 1200
Db 1141 SCRNSSTVSYLAKKEDSVSSQSSDAGLSGDSHEDTLTIDVSAISNLRKHVSARLVEDI 1200
QY 1201 GHETLYLPLEAKKEGAFVFLFHEIDRLSDLGISYGISSETTLEELFLKVAEESGVDAE 1260
Db 1201 GHETLYLPLEAKKEGAFVFLFHEIDRLSDLGISYGISSETTLEELFLKVAEESGVDAE 1260
QY 1261 TSDGTLPARNRRAFGKQSCLEPFTEDDAADPNDSIDIPESRETDLLSGMGKGSYQK 1320
Db 1261 TSDGTLPARNRRAFGKQSCLEPFTEDDAADPNDSIDIPESRETDLLSGMGKGSYQK 1320
QY 1321 GWKLTQOQFVALLMKRLLIARSRKGFPAQIYLPAAVFCIALVSLIYPPFKRPSLELO 1380
Db 1321 GWKLTQOQFVALLMKRLLIARSRKGFPAQIYLPAAVFCIALVSLIYPPFKRPSLELO 1380
QY 1381 PMWYNEGYFVSDADAPDPTGTELLNALTKDPGFRGMEGNIPPTPCQAGEEETTPAP 1440
Db 1381 PMWYNEGYFVSDADAPDPTGTELLNALTKDPGFRGMEGNIPPTPCQAGEEETTPAP 1440
QY 1441 VPQTIMDLFONGNWTQONSFPACQCSSDKIKKMLPVCPPGAGGLPPOQKQMTADILQDL 1500
Db 1441 VPQTIMDLFONGNWTQONSFPACQCSSDKIKKMLPVCPPGAGGLPPOQKQMTADILQDL 1500
QY 1501 TGNRISDYLVKTYVOIITIASLKNKIVNBERFGSGSLVSNQALPPSOEVDAIKOMK 1560
Db 1501 TGNRISDYLVKTYVOIITIASLKNKIVNBERFGSGSLVSNQALPPSOEVDAIKOMK 1560
QY 1561 HLKIAKSSADRFUNSLGRMTGLDTRNNVKWPNKKGHAISPLNVNNAILKANLOK 1620
Db 1561 HLKIAKSSADRFUNSLGRMTGLDTRNNVKWPNKKGHAISPLNVNNAILKANLOK 1620
QY 1621 GENPSHYGITAENHPLNLTQOOLSEVALMTTSYDVLVSCVIFAMSFVPASFVFLQER 1680
Db 1621 GENPSHYGITAENHPLNLTQOOLSEVALMTTSYDVLVSCVIFAMSFVPASFVFLQER 1680
QY 1681 VSKAKHLQFISGVKPYIYMLSNFVWDMCNVVPATLVIIIFCFQOKSVSSSTNLPVAL 1740
Db 1681 VSKAKHLQFISGVKPYIYMLSNFVWDMCNVVPATLVIIIFCFQOKSVSSSTNLPVAL 1740
QY 1741 LLLLYGMSITPLMYPASFVEFKIPSTAYVVLTSVNLFIGINGSVATFVLELFTDNKLNIN 1800
Db 1741 LLLLYGMSITPLMYPASFVEFKIPSTAYVVLTSVNLFIGINGSVATFVLELFTDNKLNIN 1800
QY 1801 DILKSVFLIFPHECLRGILIDWYKNQAMADALBERGENFVSPLSMDLYGRNLFMAAVEG 1860
Db 1801 DILKSVFLIFPHECLRGILIDWYKNQAMADALBERGENFVSPLSMDLYGRNLFMAAVEG 1860
QY 1861 VVFFLITVLIOYRFFLRPPVNAKLSPLNDEDEDVRRERQRLIDGGGQNDILEIKELTKI 1920
Db 1861 VVFFLITVLIOYRFFLRPPVNAKLSPLNDEDEDVRRERQRLIDGGGQNDILEIKELTKI 1920
QY 1921 YRRRRKPAVDRLICVGIIPCECGLLGVNAGKSSFTKMLTGTDTYTRGDAFLNKSIIISN 1980
Db 1921 YRRRRKPAVDRLICVGIIPCECGLLGVNAGKSSFTKMLTGTDTYTRGDAFLNKSIIISN 1980
QY 1981 IHEVHONMGYQOPDAILTELLTGREHVEFFALLKRVPEKEVKVGEAMIRKLGIVKYEK 2040
Db 1981 IHEVHONMGYQOPDAILTELLTGREHVEFFALLKRVPEKEVKVGEAMIRKLGIVKYEK 2040

QY 2041 YAGNYSGNKKRLSTAMALIGPPVVFLEDEPTTGMDKRRARFLMNCALSVYKESRSVLT 2100
 CC ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 CC cholesterol*) in the blood correlate with a high risk of cardiovascular
 CC disease, particularly coronary artery disease, but also cerebrovascular
 CC disease, coronary restenosis, and peripheral vascular disease.
 Db 2041 YAGNYSGNKKRLSTAMALIGPPVVFLEDEPTTGMDKRRARFLMNCALSVYKESRSVLT 2100
 CC ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 CC Conversely, a high level of HDL has protective effects against
 CC cardiovascular disease. The invention provides genetic constructs and
 CC transgenic cells and non-human animals comprising human ABC1 nucleic
 CC acids, and methods of gene therapy for the treatment or prevention of
 CC cardiovascular disease comprising the administration of an expression
 CC vector encoding ABC1 or an active fragment thereof. The invention also
 CC encompasses compounds which mimic ABC1 activity, compounds which
 CC stimulate ABC1 expression and methods of screening for such compounds.
 CC It further relates to methods for determining whether a patient has an
 CC increased risk for cardiovascular disease due to polymorphisms in the
 CC ABC1 gene. Human ABC1 proteins and nucleotides can be used to treat
 CC or prevent cardiovascular disease, especially coronary artery disease,
 CC cerebrovascular disease, coronary restenosis or peripheral vascular
 CC disease. They may also be used in the treatment of diseases associated
 CC with ABC1 biological activity, such as Alzheimer's disease, Niemann-Pick
 CC disease, Huntington's disease, X-linked adrenoleukodystrophy and cancer.
 CC The invention specifically excludes proteins with the exact amino acid
 CC sequences of GenBank Accession No: CAA10005.1 and X75926, and the nucleic
 CC acid with the exact sequence as GenBank Accession No: AJ012376.1. The
 CC present sequence represents the human ABC1 cholesterol transporter.
 XX
 DE Human ABC1 cholesterol transporter.
 XX
 XX Human ABC1 cholesterol transporter; chromosome 9q31;
 KW ATP-binding cassette; HDL deficiency disorder; high density lipoprotein;
 KW Tangier disease; TD; familial HDL deficiency; FHA; polymorphism;
 KW cardiovascular disease; coronary artery disease; coronary restenosis;
 KW cerebrovascular disease; peripheral vascular disease;
 KW Alzheimer's disease; Niemann-Pick disease; Huntington's disease;
 KW X-linked adrenoleukodystrophy; cancer; gene therapy; genetic diagnosis;
 KW prognosis; prophylaxis; drug screening; transgenic animal.
 XX
 OS Homo sapiens.
 XX
 PN MO200055318-A2.
 XX
 PD 21-SEP-2000.
 XX
 PF 15-MAR-2000; 2000MO-IB00532.
 XX
 PR 15-MAR-1999; 99US-0124702.
 PR 08-JUN-1999; 99US-0138048.
 PR 17-JUN-1999; 99US-0139600.
 PR 01-SEP-1999; 99US-0151977.
 XX
 PA (UYBR-) UNIV BRITISH COLUMBIA.
 PA (XENO-) XENON BIORESEARCH INC.
 XX
 PI Hayden MR, Wilson AR, Pimstone SN;
 XX
 DR WPI; 2000-587528/55.
 DR N-PSDB; AAC69120.
 XX
 PT New ABC1 polypeptide is useful for treating diseases associated with
 PT ABC1 biological activity, e.g. Alzheimer's disease, Huntington's
 PT disease and cancer -
 XX
 PS Claim 5; Page 152-157; 229pp; English.
 XX
 CC The invention relates to the human ABC1 cholesterol transporter protein
 CC (AB38082) and to nucleic acid sequences (C69120) which encode it. ABC1 is
 CC a member of the ATP-binding cassette (ABC transporter) superfamily of
 CC proteins, and plays a crucial role in cholesterol transport, particularly
 CC intracellular cholesterol trafficking in monocytes and fibroblasts, being
 CC involved in cholesterol efflux from the cell. The gene encoding ABC1 is
 CC located on chromosome 9q31, and mutations in this gene are associated
 CC with two genetic HDL (high density lipoprotein) deficiency disorders,
 CC Tangier disease (TD) and familial HDL deficiency (FHA). These diseases
 CC are distinguishable in that TD is an autosomal recessive disorder, while

CC FHA is inherited as an autosomal dominant trait. Low levels of HDL ('good
 CC cholesterol') in the blood correlate with a high risk of cardiovascular
 CC disease, particularly coronary artery disease, but also cerebrovascular
 CC disease, coronary restenosis, and peripheral vascular disease.
 CC Conversely, a high level of HDL has protective effects against
 CC cardiovascular disease. The invention provides genetic constructs and
 CC transgenic cells and non-human animals comprising human ABC1 nucleic
 CC acids, and methods of gene therapy for the treatment or prevention of
 CC cardiovascular disease comprising the administration of an expression
 CC vector encoding ABC1 or an active fragment thereof. The invention also
 CC encompasses compounds which mimic ABC1 activity, compounds which
 CC stimulate ABC1 expression and methods of screening for such compounds.
 CC It further relates to methods for determining whether a patient has an
 CC increased risk for cardiovascular disease due to polymorphisms in the
 CC ABC1 gene. Human ABC1 proteins and nucleotides can be used to treat
 CC or prevent cardiovascular disease, especially coronary artery disease,
 CC cerebrovascular disease, coronary restenosis or peripheral vascular
 CC disease. They may also be used in the treatment of diseases associated
 CC with ABC1 biological activity, such as Alzheimer's disease, Niemann-Pick
 CC disease, Huntington's disease, X-linked adrenoleukodystrophy and cancer.
 CC The invention specifically excludes proteins with the exact amino acid
 CC sequences of GenBank Accession No: CAA10005.1 and X75926, and the nucleic
 CC acid with the exact sequence as GenBank Accession No: AJ012376.1. The
 CC present sequence represents the human ABC1 cholesterol transporter.
 XX
 SQ Sequence 2261 AA:
 Query Match 99.9%; Score 11789; DB 21; Length 2261;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 2258; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MACMPQLRLMLKMLTFRROTQQLLEVAMPFLFILISRLSYRPPYDHECHFPNKA 60
 Db 1 MACMPQLRLMLKMLTFRROTQQLLEVAMPFLFILISRLSYRPPYDHECHFPNKA 60
 QY 61 MPSAGTLRWNGGIIICNANNPCFRYPTEGAPGVGNFKSVARLFGDARLLYSOKDT 120
 Db 61 MPSAGTLRWNGGIIICNANNPCFRYPTEGAPGVGNFKSVARLFGDARLLYSOKDT 120
 QY 121 SMKDMRKVLRLLQQLIKSSSNLKLDDFLVDNETFSGFLYHNLSPKSTVDKMLRADYILH 180
 Db 121 SMKDMRKVLRLLQQLIKSSSNLKLDDFLVDNETFSGFLYHNLSPKSTVDKMLRADYILH 180
 QY 181 KVFLOGYQLHLTSLCNGSKSEEMIDLGQEVSELCGLPKKELAAERYLRNMDILPIL 240
 Db 181 KVFLOGYQLHLTSLCNGSKSEEMIDLGQEVSELCGLPREKLAERYLRNMDILPIL 240
 QY 241 RRLNSTSPPEPKELAEATKTLTSLGTLAOLFFSMRSWMDKROEYMLFTVNNSSSSSTOI 300
 Db 241 RRLNSTSPPEPKELAEATKTLTSLGTLAOLFFSMRSWMDKROEYMLFTVNNSSSSSTOI 300
 QY 301 YQAVSRIVCGHEPGGGLIKISLWYEDNNYKALFEGNGSTEDDAETFYDNSTTPPCNDLAK 360
 Db 301 YQAVSRIVCGHEPGGGLIKISLWYEDNNYKALFEGNGSTEDDAETFYDNSTTPPCNDLAK 360
 QY 361 NLESPPLSLRIIMKALKPLLVGKLLTTPDTPATROYMAEVNKPFOELAEFHLDGMEELS 420
 Db 361 NLESPPLSLRIIMKALKPLLVGKLLTTPDTPATROYMAEVNKPFOELAEFHLDGMEELS 420
 QY 421 PKIWFPMENSQEMDLVRLMLDSRDNDHFEQDLGIDMTADIVAFIAKHPEDVOSSNGS 480
 Db 421 PKIWFPMENSQEMDLVRLMLDSRDNDHFEQDLGIDMTADIVAFIAKHPEDVOSSNGS 480
 QY 481 VYTWREAFNETNOAIRTISRFEVCVNLKLEPLATEVWLINSSMLLDERKRFAGIVTFG 540
 Db 481 VYTWREAFNETNOAIRTISRFEVCVNLKLEPLATEVWLINSSMLLDERKRFAGIVTFG 540
 QY 541 ITPGSELEPHHYKTRMDIDNVERTNKTKDGYWDPGPADEPDMRYWVGFAVLQDGV 600
 Db 541 ITPGSELEPHHYKTRMDIDNVERTNKTKDGYWDPGPADEPDMRYWVGFAVLQDGV 600
 QY 601 EDAIIRVLGTGTEKRTGVYQOMPVCYVDIDILRVWSRMPFLPTLAWIYSAVVIKGIY 660
 Db 601 EDAIIRVLGTGTEKRTGVYQOMPVCYVDIDILRVWSRMPFLPTLAWIYSAVVIKGIY 660

```

Db 601 EQAIIIRVLGTEKKTGYVMQMPYVVDITFLRVMSRSMPLFMTLAMIYSVAVIIKGIY 660
QY 661 YEKEARKETMRIMGLDINSILWFSEWISLIPLVASALLVILIKLGNLIPYSPVAVF 720
Db 661 YEKEARKETMRIMGLDINSILWFSEWISLIPLVASALLVILIKLGNLIPYSPVAVF 720
QY 721 FLISVAVVITLLOCFILSTLFSRANLAACGGIIYFTLIPVLCVAMODVYGFLLKIFAS 780
Db 721 FLISVAVVITLLOCFILSTLFSRANLAACGGIIYFTLIPVLCVAMODVYGFLLKIFAS 780
QY 781 LLSPAFGGCFEYALFEEOGIVQWMDLFSPEDEBNITTSISMLEDITLYGVMY 840
Db 781 LLSPAFGGCFEYALFEEOGIVQWMDLFSPEDEBNITTSISMLEDITLYGVMY 840
QY 841 YLEAVEFPOYGIIPRWPYEPCTKSYWFGESDEKSHPGSNOKRMEICMEEPHILKIGVS 900
Db 841 YLEAVEFPOYGIIPRWPYEPCTKSYWFGESDEKSHPGSNOKRMEICMEEPHILKIGVS 900
QY 901 IONLVKVRDGMKAVVDGLALNFEEOITTSFLGHNGAGKTTMSILNGLFPPTSGTAYIL 960
Db 901 IONLVKVRDGMKAVVDGLALNFEEOITTSFLGHNGAGKTTMSILNGLFPPTSGTAYIL 960
QY 961 GKDISEKSTIRONLGVCPHNNVLFDMITVEHITWFAIRLKGISEKIVAKEMQALDVG 1020
Db 961 GKDISEKSTIRONLGVCPHNNVLFDMITVEHITWFAIRLKGISEKIVAKEMQALDVG 1020
QY 1021 LPSKSLKSTISQLSGMMOKRLSVALAFVGSKVYILDEPTAGVDPYSRRKTIWELLKRYO 1080
Db 1021 LPSKSLKSTISQLSGMMOKRLSVALAFVGSKVYILDEPTAGVDPYSRRKTIWELLKRYO 1080
QY 1081 GRTIITLHHMDEADVIGDRIATISHRKLCQVSSLEFLKMOLOTGYLTLVKKDVSSIS 1140
Db 1081 GRTIITLHHMDEADVIGDRIATISHRKLCQVSSLEFLKMOLOTGYLTLVKKDVSSIS 1140
QY 1141 SCRNSSTSVYSLKEDSVSSSDAGLSDHESDITLIDVSATSNLIRKHVSARLVEI 1200
Db 1141 SCRNSSTSVYSLKEDSVSSSDAGLSDHESDITLIDVSATSNLIRKHVSARLVEI 1200
QY 1201 GHELTLYLPEYAAKEGAFVELFHEIDRLSDLGISSTLLEIFLKAEEGVDAE 1260
Db 1201 GHELTLYLPEYAAKEGAFVELFHEIDRLSDLGISSTLLEIFLKAEEGVDAE 1260
QY 1261 TSDGTLPARRRRRAFGDKOSCLRPFTEDDAADPNDSDIDPESRETDLSGMGKGSYQVK 1320
Db 1261 TSDGTLPARRRRRAFGDKOSCLRPFTEDDAADPNDSDIDPESRETDLSGMGKGSYQVK 1320
QY 1321 GMLTLOOQFVALLMKRLLIARRSRKGFQAQIYLPVAVFCIALVFSLIVPFCKYPSLEO 1380
Db 1321 GMLTLOOQFVALLMKRLLIARRSRKGFQAQIYLPVAVFCIALVFSLIVPFCKYPSLEO 1380
QY 1381 PWMYNEGYTVSNDAPEDTGTLELINALTKDPCGTRCMEGNIPRTPCQAGEEWTAP 1440
Db 1381 PWMYNEGYTVSNDAPEDTGTLELINALTKDPCGTRCMEGNIPRTPCQAGEEWTAP 1440
QY 1441 VPQTIMDLFONGMWTQNPSPACQSSDKIKKMLPVCPPGAGGLPPRQKONTADILDDL 1500
Db 1441 VPQTIMDLFONGMWTQNPSPACQSSDKIKKMLPVCPPGAGGLPPRQKONTADILDDL 1500
QY 1501 TGRNISYLYKTYVQIIAASLKNKIVNEFRYGFSLGVSNQALPQOEVDATKQKK 1560
Db 1501 TGRNISYLYKTYVQIIAASLKNKIVNEFRYGFSLGVSNQALPQOEVDATKQKK 1560
QY 1561 HLKLAQSSADRFNLNGLRFMTGLDTRNNVKKWPNKKGHAISFLANTINNALIRANLOK 1620
Db 1561 HLKLAQSSADRFNLNGLRFMTGLDTRNNVKKWPNKKGHAISFLANTINNALIRANLOK 1620
QY 1621 GENPSSHGITAENHPLNLTQOOLSEVALMTTSDVLVSCIVAFAMSEFPAFVFLIOER 1680
Db 1621 GENPSSHGITAENHPLNLTQOOLSEVALMTTSDVLVSCIVAFAMSEFPAFVFLIOER 1680
QY 1681 VSKAKHLOFISGVKPIYIWLSNFVDMCNVYVPATLVIITIFICHQOKSYSSSTMPLVAL 1740
Db 1681 VSKAKHLOFISGVKPIYIWLSNFVDMCNVYVPATLVIITIFICHQOKSYSSSTMPLVAL 1740

```

```

QY 1741 LLLLVGMSITPLMPASFVEKIPSTAVVITLTVNLFTIGINGSVATFVLELTDNKLNNIN 1800
Db 1741 LLLLVGMSITPLMPASFVEKIPSTAVVITLTVNLFTIGINGSVATFVLELTDNKLNNIN 1800
QY 1801 DILKSVFLIPFHCLGRLLIDMVKNQAMADALEFGENRRVSPLSMDLVGRNLFAMAVBG 1860
Db 1801 DILKSVFLIPFHCLGRLLIDMVKNQAMADALEFGENRRVSPLSMDLVGRNLFAMAVBG 1860
QY 1861 VVEFLITVLIOYRFFIPRRPNAKLSPLNDEDEVDREORILDDGGONDILEIKELTETI 1920
Db 1861 VVEFLITVLIOYRFFIPRRPNAKLSPLNDEDEVDREORILDDGGONDILEIKELTETI 1920
QY 1921 YRRKRRKPVADICVGIPEGECFGLLVGNAGKSSFKMLTGDITTVTRGDAFLINKSILSN 1980
Db 1921 YRRKRRKPVADICVGIPEGECFGLLVGNAGKSSFKMLTGDITTVTRGDAFLINKSILSN 1980
QY 1981 IHEYHONNGYCPQPDALTTELJGHEHEFFALLRGVPERKVGVEGMAIRKILGVKYGK 2040
Db 1981 IHEYHONNGYCPQPDALTTELJGHEHEFFALLRGVPERKVGVEGMAIRKILGVKYGK 2040
QY 2041 YAGNYSGGNKKRLSTANALIGCPPVFLDEPTTGMPKARFELMNCALSVKEGSRVLT 2100
Db 2041 YAGNYSGGNKKRLSTANALIGCPPVFLDEPTTGMPKARFELMNCALSVKEGSRVLT 2100
QY 2101 SHSMECEALCTRAIIVNCRFRCLGSVQHLKRRFGDITTVIRIAGSNPDLKPVQDEFG 2160
Db 2101 SHSMECEALCTRAIIVNCRFRCLGSVQHLKRRFGDITTVIRIAGSNPDLKPVQDEFG 2160
QY 2161 LAFPGSVLKEKRNMLQYQPSLSSILARLFTLSQSKRLHIEBYSVQOTLDDQVFVNF 2220
Db 2161 LAFPGSVLKEKRNMLQYQPSLSSILARLFTLSQSKRLHIEBYSVQOTLDDQVFVNF 2220
QY 2221 AKQSDDDHLKDLSLHKNQTVADVAVLTSFLODEKYESYV 2261
Db 2221 AKQSDDDHLKDLSLHKNQTVADVAVLTSFLODEKYESYV 2261

RESULT 8
AAB71749
ID AAB71749 standard; protein; 2261 AA.
XX
AC AAB71749;
XX
DT 17-MAY-2001 (first entry)
XX
DE Human ABC1 protein.
XX
KW High density lipoprotein-cholesterol; HDL-C; cardiovascular; ABC1.
XX
OS Homo sapiens.
XX
PN W0200115676-A2.
XX
PD 08-MAR-2001.
XX
PF 01-SEP-2000; 2000WO-1B01492.
XX
PR 01-SEP-1999; 99US-0151977.
XX
PR 15-MAR-2000; 2000US-0526193.
XX
PR 23-JUN-2000; 2000US-0213958.
XX
PA (UYBR-) UNIV BRITISH COLUMBIA.
XX
PA (XENO-) XENON GENETICS INC.
XX
PI Hayden MR, Brooks-Wilson AR, Pimstone SN, Clee SM;
XX
DR WPL; 2001-244356/25.
XX
PT
PT Treating a lower than normal high density lipoprotein-cholesterol
PT (HDL-C) level, a higher than normal triglyceride level, or a
PT cardiovascular disease, by administering a compound that modulates LXR-
PT or RXR-mediated transcriptional activity -

```

XX Claim 16; Fig 2; 317pp; English.

CC The present invention relates to a method for treating a patient
CC diagnosed as having a lower than normal high density
CC lipoprotein-cholesterol (HDL-C) level, a higher than normal
CC triglyceride level, or a cardiovascular disease, involving
CC administering a compound that modulates LXR- or RXR-mediated
CC transcriptional activity or ABC1 expression or activity.
CC The LXR gene product may be used in an assay to identify
CC compounds useful for the treatment of a disease or condition selected a
CC lower than normal HDL cholesterol level, a higher than normal
CC triglyceride level, and a cardiovascular disease.

XX Sequence 2261 AA;

Query Match 99.9%; Score 11789; DB 22; Length 2261;

Best Local Similarity 99.9%; Pred. No. 0;

Matches 2258; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 MACWPQLRLMKNTFRRCOTCOLLEFVAMPLEFILLISVRLSPYEQHECHPDKA 60
DB 1 MACWPQLRLMKNTFRRCOTCOLLEFVAMPLEFILLISVRLSPYEQHECHPDKA 60
QY 61 MESAGTLPWVGILICANNPCFRYPTEGAPGVGNFNKSIARLFSDARRLTYSDKT 120
DB 61 MESAGTLPWVGILICANNPCFRYPTEGAPGVGNFNKSIARLFSDARRLTYSDKT 120
QY 121 SKMDKRVLRITLQIKKSSSNLKLODFLVNDNETFSGFLYHNLSPKSYVDKMLRADVILH 180
DB 121 SKMDKRVLRITLQIKKSSSNLKLODFLVNDNETFSGFLYHNLSPKSYVDKMLRADVILH 180
QY 181 KYVLOGLYOLHLSLNGSGSEEMTOLGDOVESELCGLPEKLAARVYRSMDITKPL 240
DB 181 KYVLOGLYOLHLSLNGSGSEEMTOLGDOVESELCGLPEKLAARVYRSMDITKPL 240
QY 241 RFLNSTSPPEKELATATLTLHSLCTLAQELFMSRMSDMROEVLFLNVSSSSSTOI 300
DB 241 RFLNSTSPPEKELATATLTLHSLCTLAQELFMSRMSDMROEVLFLNVSSSSSTOI 300
QY 301 YQAVSTIVCGHPREGGLIKSLNMYEDNNYKALFGNGTEDEAETFYDNTSTPYCNDLTK 360
DB 301 YQAVSTIVCGHPREGGLIKSLNMYEDNNYKALFGNGTEDEAETFYDNTSTPYCNDLTK 360
QY 361 NLESSPLSRITKALPPLVYGLTYPPDPATROVAENKTPFOELAVHDELGMEELS 420
DB 361 NLESSPLSRITKALPPLVYGLTYPPDPATROVAENKTPFOELAVHDELGMEELS 420
QY 421 PKIWMENSOEMDLVRLMLDSRDNDHFWEQQLDGLDWTADIVAFIAKHPEDVOSSNGS 480
DB 421 PKIWMENSOEMDLVRLMLDSRDNDHFWEQQLDGLDWTADIVAFIAKHPEDVOSSNGS 480
QY 481 VYTWREAFNETNOAIRTISRFEVCVNLKLEPATLATEVWLINKSMELDERKFPAGIVFTG 540
DB 481 VYTWREAFNETNOAIRTISRFEVCVNLKLEPATLATEVWLINKSMELDERKFPAGIVFTG 540
QY 541 ITTGESIEPLPHVYKIRMDIDNERTNKIKDGYWDPGRADPEDRMRYWGGATYQDYY 600
DB 541 ITTGESIEPLPHVYKIRMDIDNERTNKIKDGYWDPGRADPEDRMRYWGGATYQDYY 600
QY 601 EOAIIIRVLCTEKKTGVMQOMPYPCVYDDIPLRVMSRMPLEMTLAWISVAIIKGIY 660
DB 601 EOAIIIRVLCTEKKTGVMQOMPYPCVYDDIPLRVMSRMPLEMTLAWISVAIIKGIY 660
QY 661 YEKFAKLEKMRIMGLDNLILMFSSILPILVSGGLVYLKGNLLPYSDPSVYV 720
DB 661 YEKFAKLEKMRIMGLDNLILMFSSILPILVSGGLVYLKGNLLPYSDPSVYV 720
QY 721 FLISFAVVTLLQGLFISTESSRANLAAACGIIYFTLYPLVYLCVAMODVGTIKTFAS 780
DB 721 FLISFAVVTLLQGLFISTESSRANLAAACGIIYFTLYPLVYLCVAMODVGTIKTFAS 780
QY 781 LLSVAFGCEYFALFEEOGIGVOMDNLFESPVEEDGFNLJTSSIMLFDTEFLYGVMTW 840

DB 781 LLSVAFGCEYFALFEEOGIGVOMDNLFESPVEEDGFNLJTSSIMLFDTEFLYGVMTW 840
QY 841 YIEAFVPGGYGIPRPWPFPCTKSWNCEESDEKSHGSSNOKRMSEICMEEPHLLKIGVS 900
DB 841 YIEAFVPGGYGIPRPWPFPCTKSWNCEESDEKSHGSSNOKRMSEICMEEPHLLKIGVS 900
QY 901 IONLVKRYRGMKAVYDGLALNFEGOITSFLHNGAGKTTMSILTGLFPPTSGTAYIL 960
DB 901 IONLVKRYRGMKAVYDGLALNFEGOITSFLHNGAGKTTMSILTGLFPPTSGTAYIL 960
QY 961 GKDIRSEMSTIRONLVGCPQHNVLFDMLTVEEHIWFAYLKGLSEKHVAKEMOMALDVG 1020
DB 961 GKDIRSEMSTIRONLVGCPQHNVLFDMLTVEEHIWFAYLKGLSEKHVAKEMOMALDVG 1020
QY 1021 LPSSKLKSKTSQSLSGMORKLSVALAFVGSKYVILDEPAGDPSRRGIMELLKRYR 1080
DB 1021 LPSSKLKSKTSQSLSGMORKLSVALAFVGSKYVILDEPAGDPSRRGIMELLKRYR 1080
QY 1081 GRTIILSTHMDADVLDRIATISHGKLCVSSSFLKNOGTGYVLTLYKRDVSSLS 1140
DB 1081 GRTIILSTHMDADVLDRIATISHGKLCVSSSFLKNOGTGYVLTLYKRDVSSLS 1140
QY 1141 SCRNSSTVSYLKKEDSVSSDAGLGSDESPDTLTDVSAISNLRKXVSEARLYEDI 1200
DB 1141 SCRNSSTVSYLKKEDSVSSDAGLGSDESPDTLTDVSAISNLRKXVSEARLYEDI 1200
QY 1201 GHELTLYVPEAKKGAFAVELFHEIDRLSDLGTSISGISTTLEELFLVAEESGDVAE 1260
DB 1201 GHELTLYVPEAKKGAFAVELFHEIDRLSDLGTSISGISTTLEELFLVAEESGDVAE 1260
QY 1261 TSDGTLPARNRRAFGDKOSCLRPFTEDDADPNDSDIDPESRETDLISGMDGKSYQV 1320
DB 1261 TSDGTLPARNRRAFGDKOSCLRPFTEDDADPNDSDIDPESRETDLISGMDGKSYQV 1320
QY 1321 GSKLTQOQFVALLMKRLLIARRSKKGFPAQVLPAYVYCALVSLVPPGKTPSLEQ 1380
DB 1321 GSKLTQOQFVALLMKRLLIARRSKKGFPAQVLPAYVYCALVSLVPPGKTPSLEQ 1380
QY 1381 PMVNEOYTFVSNDAPEDTGTLELINALTKDPGFGTCMEGNPJPDPICQAGEEMTAP 1440
DB 1381 PMVNEOYTFVSNDAPEDTGTLELINALTKDPGFGTCMEGNPJPDPICQAGEEMTAP 1440
QY 1441 VPQTTIMDLFONGNMTQNPSPACQSSDKIKMLPVCPGAGGLPPQKONTADILQDL 1500
DB 1441 VPQTTIMDLFONGNMTQNPSPACQSSDKIKMLPVCPGAGGLPPQKONTADILQDL 1500
QY 1501 TGRNISDYLVTYVOITAKSLKNKIWNEFPYGGFSLGVSTQALPSPQEVNDAIKQMK 1560
DB 1501 TGRNISDYLVTYVOITAKSLKNKIWNEFPYGGFSLGVSTQALPSPQEVNDAIKQMK 1560
QY 1561 HUKLAKSSADREFLNSLGRPMTGLDTRNNVAVFNNKGMHAISFLANINNAIRANLQK 1620
DB 1561 HUKLAKSSADREFLNSLGRPMTGLDTRNNVAVFNNKGMHAISFLANINNAIRANLQK 1620
QY 1621 GENPSHYGTTAFNHLNLTQOULSEVALMTTSVDVLSICVIFAMSEVPASVVFLLQER 1680
DB 1621 GENPSHYGTTAFNHLNLTQOULSEVALMTTSVDVLSICVIFAMSEVPASVVFLLQER 1680
QY 1681 VSKAKHLOFISGVKRYVYWLNFVNDKCNVYVPAVLVILITFCQOKSYVSTNLPVAL 1740
DB 1681 VSKAKHLOFISGVKRYVYWLNFVNDKCNVYVPAVLVILITFCQOKSYVSTNLPVAL 1740
QY 1741 LLLLGMSITPLMPYASPFVKIPSTAYVVLTSVNLFIGINSVAFVLELTDKLNIN 1800
DB 1741 LLLLGMSITPLMPYASPFVKIPSTAYVVLTSVNLFIGINSVAFVLELTDKLNIN 1800
QY 1801 DILKSVLLFPFHCGLRGILIDVKNQAMADALERFGENRFPVSLWLVGRNLVMAVEG 1860
DB 1801 DILKSVLLFPFHCGLRGILIDVKNQAMADALERFGENRFPVSLWLVGRNLVMAVEG 1860
QY 1861 VVEFLITVLIQYRFRFPVNAKLSPLNDEDEVDVRRRQRTLDGSGGNDILEIKELKI 1920
DB 1861 VVEFLITVLIQYRFRFPVNAKLSPLNDEDEVDVRRRQRTLDGSGGNDILEIKELKI 1920

Db 1861 VVFELTLVLYRFFIRPRPVNAKLSPUNDEDEYRREKQRTLDGGGNDILEIKELTKI 1920
 QY 1921 YRRKKRPADVRLTCVIGPEGCEGGLGVNGAGKSTFKMLTGDPTVTNRGDAFLNKNSILSN 1980
 Db 1921 YRRKKRPADVRLTCVIGPEGCEGGLGVNGAGKSTFKMLTGDPTVTNRGDAFLNKNSILSN 1980
 QY 1981 HEHVHNMGYCOPAFITELLGRHVEFFALLRCVPEKEKVGEMAIRKLGIVKYEK 2040
 Db 1981 HEHVHNMGYCOPAFITELLGRHVEFFALLRCVPEKEKVGEMAIRKLGIVKYEK 2040
 QY 2041 YAGNVSNGKRRKLTAMALIGGPVVFDEPTTGMDPKARRFLNMCALSYVKEGRSVLT 2100
 Db 2041 YAGNVSNGKRRKLTAMALIGGPVVFDEPTTGMDPKARRFLNMCALSYVKEGRSVLT 2100
 QY 2101 SHSMECEALCTRMAIMVNGRRCLGSVOHLKNRFGDYTTIVRIAGSNPDLKPVQDFPG 2160
 Db 2101 SHSMECEALCTRMAIMVNGRRCLGSVOHLKNRFGDYTTIVRIAGSNPDLKPVQDFPG 2160
 QY 2161 IAFPGSVLKEKRRMLQYQPLSSLSLARIFSLISQSKRRLHIEDYSQTTLDQVEYNF 2220
 Db 2161 IAFPGSVLKEKRRMLQYQPLSSLSLARIFSLISQSKRRLHIEDYSQTTLDQVEYNF 2220
 QY 2221 AKDQSDDDLKDLSLHKNOTVVDVAVLTSFLQDEKVKESYV 2261
 Db 2221 AKDQSDDDLKDLSLHKNOTVVDVAVLTSFLQDEKVKESYV 2261
 RESULT 9
 AAB31363
 ID AAB31363 standard; Protein: 2261 AA.
 AC AAB31363;
 DT 20-Apr-2001 (first entry)
 XX
 DE Amino acid sequence of ABC1 polypeptide from Tangier disease patient.
 KW Human: adenosine triphosphate binding cassette protein 1; ABC1;
 KW apolipoprotein-mediated mobilisation; cholesterol; Tangier disease;
 KW chromosome 9q22-9q31; heart disease; hypercholesterolemia;
 KW atherosclerosis; cholesterol transport.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT Misc-difference 587 /note="this is changed from Arg to Trp in Tangier
 FT disease"
 PN WO200078972-A2.
 XX 28-DEC-2000.
 PD 16-JUN-2000; 2000WO-US16765.
 PF 18-JUN-1999; 9905-0140264.
 PR 14-SEP-1999; 9905-0153872.
 PR 19-NOV-1999; 9905-016573.
 XX (CVTH-) CV THERAPEUTICS INC.
 PA Lawn RM, Wade D, Garvin M;
 PI WPT; 2001-137812/14.
 DR
 XX Adenosine triphosphate (ATP) binding cassette (ABC) polynucleotide,
 PT useful for the development of agents for the treatment of heart disease
 PT and other disorders associated with hypercholesterolemia and
 PT atherosclerosis -
 XX
 PS Disclosure: Page 176-191; 215pp; English.
 XX The present sequence represents a human adenosine triphosphate (ATP)

CC binding cassette protein (ABC) 1 polypeptide, and is isolated from
 CC a Tangier disease patient. ABC1 resides in cell membranes and utilises
 CC ATP hydrolysis to transport a wide variety of substrates across the
 CC plasma membrane. ABC1 is a pivotal protein in the apolipoprotein-mediated
 CC mobilisation of intracellular cholesterol stores. ABC1 is defective in
 CC Tangier disease, a genetic disorder characterised by abnormal
 CC HDL-cholesterol metabolism. The ABC1 gene is localised to chromosome
 CC 9q22-9q31. The ABC1 genes and proteins are useful for developing
 CC pharmaceutical agents for the treatment of heart disease and other
 CC disorders associated with hypercholesterolemia and atherosclerosis. The
 CC genes are useful for developing screening assays to screen for compounds
 CC that regulate the expression of genes associated with cholesterol
 CC transport. The genes and proteins are also useful for cholesterol
 CC as diagnostic indicators of cardiovascular disease and other disorders
 CC associated with hypercholesterolemia.
 CC
 XX
 SO Sequence 2261 AA:
 Query Match 99.9%; Score 11789; DB 22; Length 2261;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2260; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 MACWPQLRLIMKNTLFRKQTCQLLEVAWPLFTLLISVRLSPYEPYEGHCFPNKA 60
 Db 1 MACWPQLRLIMKNTLFRKQTCQLLEVAWPLFTLLISVRLSPYEPYEGHCFPNKA 60
 QY 61 MPSAGTLPMVOGITCANNPCFRYPGEPGVGVGNFNSTIARLFSDARRLLLSQKDT 120
 Db 61 MPSAGTLPMVOGITCANNPCFRYPGEPGVGVGNFNSTIARLFSDARRLLLSQKDT 120
 QY 121 SMKDMKRVLTLOQIKKSSSNLKLQDFLVNDETFSGLFVHNLSPKSTYDKMLRADVILH 180
 Db 121 SMKDMKRVLTLOQIKKSSSNLKLQDFLVNDETFSGLFVHNLSPKSTYDKMLRADVILH 180
 QY 181 KYFLOGYQLHLTSLCNGSSEEMIQDQEVSELGLPEKEXILAAERYLRSMNMLKPTL 240
 Db 181 KYFLOGYQLHLTSLCNGSSEEMIQDQEVSELGLPEKEXILAAERYLRSMNMLKPTL 240
 QY 241 RITNSTSPPEKLEATPTLLHSIGTLAQELFSKRSMDMQEWPFLTNSSSSSTOI 300
 Db 241 RITNSTSPPEKLEATPTLLHSIGTLAQELFSKRSMDMQEWPFLTNSSSSSTOI 300
 QY 301 YQAVSRIVCGHEPGEGLIKSLNMYEDNNYKALFPGNGTEDEAETFYDNSTPYCNDLAK 360
 Db 301 YQAVSRIVCGHEPGEGLIKSLNMYEDNNYKALFPGNGTEDEAETFYDNSTPYCNDLAK 360
 QY 361 NLESSPLSLIITKALKPLLVGKILYPTPTPATROYMAEVNKTFOELAVFHDLGEMWELIS 420
 Db 361 NLESSPLSLIITKALKPLLVGKILYPTPTPATROYMAEVNKTFOELAVFHDLGEMWELIS 420
 QY 421 PKIWTFMENSQEMDLVRMLLDSRDNDFHEQQLDGLMTADIDYAFIAKHPEDVOSSNGS 480
 Db 421 PKIWTFMENSQEMDLVRMLLDSRDNDFHEQQLDGLMTADIDYAFIAKHPEDVOSSNGS 480
 QY 481 VYTWBPAFNETNOAIRTISREMECVNLKLEPIATEVMILINKSMELDERKFMAGIYFTG 540
 Db 481 VYTWBPAFNETNOAIRTISREMECVNLKLEPIATEVMILINKSMELDERKFMAGIYFTG 540
 QY 541 ITPGSEILPHNKKYKIRNDIDNVEFTNKIKQGYWDCGRADPFEDMYVMGCAFYLDVV 600
 Db 541 ITPGSEILPHNKKYKIRNDIDNVEFTNKIKQGYWDCGRADPFEDMYVMGCAFYLDVV 600
 QY 601 EQAITRVLTGTEKKTGYVMQMPYCYVDJFLRWMSRMLPMTLAMIYSVAIITIGIV 660
 Db 601 EQAITRVLTGTEKKTGYVMQMPYCYVDJFLRWMSRMLPMTLAMIYSVAIITIGIV 660
 QY 661 YEKEARKLETKRIMGLDLSILFMSWISSILPLIVSAGLLVILKGLNLPLSPSPSVFV 720
 Db 661 YEKEARKLETKRIMGLDLSILFMSWISSILPLIVSAGLLVILKGLNLPLSPSPSVFV 720
 QY 721 FLSTAVAVVTIIQCFILSTLFSRANLAACGIIFTLYLPVLCVAMQDVYGFLLKIFAS 780
 Db 721 FLSTAVAVVTIIQCFILSTLFSRANLAACGIIFTLYLPVLCVAMQDVYGFLLKIFAS 780

```

QY 781 LLSPAFGGCEYFALFEEOGIGVOMDNLFESEVEDGFLNLTTSISMMLFDTLYGVMTW 840
    |||||
Db 781 LLSPAFGGCEYFALFEEOGIGVOMDNLFESEVEDGFLNLTTSISMMLFDTLYGVMTW 840
QY 841 YLEAVFPGQYGTIPRWYFPCTKSYNFGESDEKSHPGSNOKRMSTICMEEPHTLKCVS 900
    |||||
Db 841 YLEAVFPGQYGTIPRWYFPCTKSYNFGESDEKSHPGSNOKRMSTICMEEPHTLKCVS 900
QY 901 IONLVKVRDGGKVAVDGALALNFYEGQITSPFHNGACKTMTSLTGLPPTSGTATIL 960
    |||||
Db 901 IONLVKVRDGGKVAVDGALALNFYEGQITSPFHNGACKTMTSLTGLPPTSGTATIL 960
QY 961 GKDISEMSTIRQNLGVCPOHNVLFDMLTVEEHIMFARLKGISEKHAKEMEQMALDVG 1020
    |||||
Db 961 GKDISEMSTIRQNLGVCPOHNVLFDMLTVEEHIMFARLKGISEKHAKEMEQMALDVG 1020
QY 1021 LPSSKLKSTSQLSGGMOKLSVALAFVGGSKVVLDEPTAGVDPYSRGIMWELLKRYQ 1080
    |||||
Db 1021 LPSSKLKSTSQLSGGMOKLSVALAFVGGSKVVLDEPTAGVDPYSRGIMWELLKRYQ 1080
QY 1081 GRTIILSTHMEADYVLDGRIATISHGLCCVGSFLPKNOLGTGYLTLYKKDVESLS 1140
    |||||
Db 1081 GRTIILSTHMEADYVLDGRIATISHGLCCVGSFLPKNOLGTGYLTLYKKDVESLS 1140
QY 1141 SCRNSSSTVYLKEDSVSSSSDAGLSDHESDITLIDVSAISMLIRKHVSEARLVEDI 1200
    |||||
Db 1141 SCRNSSSTVYLKEDSVSSSSDAGLSDHESDITLIDVSAISMLIRKHVSEARLVEDI 1200
QY 1201 GHELYTVLPYEAKGAGAYELFHEIDRLSDLGSSYGISETTLEEILFKAESGVDAE 1260
    |||||
Db 1201 GHELYTVLPYEAKGAGAYELFHEIDRLSDLGSSYGISETTLEEILFKAESGVDAE 1260
QY 1261 TSDGTLPARNRARAGDSCLRPTEDDAADPNDSIDPESREFDLISGMDGKSYQV 1320
    |||||
Db 1261 TSDGTLPARNRARAGDSCLRPTEDDAADPNDSIDPESREFDLISGMDGKSYQV 1320
QY 1321 GSKLTLQOQFVALLMKRLLIARSRKGFPAQIYLPVAFVCIALVESLIYPPGKYSELEQ 1380
    |||||
Db 1321 GSKLTLQOQFVALLMKRLLIARSRKGFPAQIYLPVAFVCIALVESLIYPPGKYSELEQ 1380
QY 1381 PMYNEQYTFVSNDAPEDTGTELLNALTKDPGFGSTRMEGNIPIPTPCQAGEEWTAP 1440
    |||||
Db 1381 PMYNEQYTFVSNDAPEDTGTELLNALTKDPGFGSTRMEGNIPIPTPCQAGEEWTAP 1440
QY 1441 VPQITMDLFPONGNMTMNPSPACQCSSDKIKMLPVCPRGAGLPBPQRKONTADIIODL 1500
    |||||
Db 1441 VPQITMDLFPONGNMTMNPSPACQCSSDKIKMLPVCPRGAGLPBPQRKONTADIIODL 1500
QY 1501 TGRNIDSLVKTYYQIIAKSLKNNKIMVNEFRYGFSLGVSNTQALPSEQEVDATIKQMK 1560
    |||||
Db 1501 TGRNIDSLVKTYYQIIAKSLKNNKIMVNEFRYGFSLGVSNTQALPSEQEVDATIKQMK 1560
QY 1561 HLKLAQSSADRFNLISLGRFMTGLDTRNNVKYVFNKKMHAISFLNVTINAILERANLQK 1620
    |||||
Db 1561 HLKLAQSSADRFNLISLGRFMTGLDTRNNVKYVFNKKMHAISFLNVTINAILERANLQK 1620
QY 1621 GENPSHYGTTAFNHPNLNTKQOLSEVALMTTSVDLVASICVIFAMSEVPASVAVFLQER 1680
    |||||
Db 1621 GENPSHYGTTAFNHPNLNTKQOLSEVALMTTSVDLVASICVIFAMSEVPASVAVFLQER 1680
QY 1681 VSKAKHLQFISGVKPVYIWLNSFVMDCNYYVPATLVIIIFCFQOKSYVSTNLPVAL 1740
    |||||
Db 1681 VSKAKHLQFISGVKPVYIWLNSFVMDCNYYVPATLVIIIFCFQOKSYVSTNLPVAL 1740
QY 1741 LLLLYGKSTPLMTASTVETKIPSTAYIVVLTSVNLFTIINSVATFVLELFTDKNNIN 1800
    |||||
Db 1741 LLLLYGKSTPLMTASTVETKIPSTAYIVVLTSVNLFTIINSVATFVLELFTDKNNIN 1800
QY 1801 DILKSVFLFPFECGRGLIDVKNQAMADALERGENRFPSPSMDVGNLPMAYEG 1860
    |||||
Db 1801 DILKSVFLFPFECGRGLIDVKNQAMADALERGENRFPSPSMDVGNLPMAYEG 1860

```

```

QY 1861 VFPLITVLIOYREFIRPRPVNAKLSPANDEDEVRERORILIDGGGQNDILEIKELTKI 1920
    |||||
Db 1861 VFPLITVLIOYREFIRPRPVNAKLSPANDEDEVRERORILIDGGGQNDILEIKELTKI 1920
QY 1921 YRRKRPAVDRIYGVIPGECFGLLVNGAGKSTKMLTGTFTVTRGDAFLNKNSILSN 1980
    |||||
Db 1921 YRRKRPAVDRIYGVIPGECFGLLVNGAGKSTKMLTGTFTVTRGDAFLNKNSILSN 1980
QY 1981 IHEYHOMMGVCPQPDATITELLTGEHEVEFALLRGVREKVGVEGMAIRKLGIVYCKE 2040
    |||||
Db 1981 IHEYHOMMGVCPQPDATITELLTGEHEVEFALLRGVREKVGVEGMAIRKLGIVYCKE 2040
QY 2041 YAGNYSOGNKRKLSTAALLIGPPVVFLEDEPTTGMPKARFPLMNCALSVKEGRSVLT 2100
    |||||
Db 2041 YAGNYSOGNKRKLSTAALLIGPPVVFLEDEPTTGMPKARFPLMNCALSVKEGRSVLT 2100
QY 2101 SHSMECEALCTRMATVNGRFRCLGVQHLKNRFGDGYIIVRIAGSNPDLKPVODFFG 2160
    |||||
Db 2101 SHSMECEALCTRMATVNGRFRCLGVQHLKNRFGDGYIIVRIAGSNPDLKPVODFFG 2160
QY 2161 LAFPGSVLKEKRRMLOYQLPSSLSLARLFSTISQSKRLIEDVSVSQTILDQVFAN 2220
    |||||
Db 2161 LAFPGSVLKEKRRMLOYQLPSSLSLARLFSTISQSKRLIEDVSVSQTILDQVFAN 2220
QY 2221 AKDQSDDDLKDLSLKRNQTVVDVAVLTSPLODEKYESYV 2261
    |||||
Db 2221 AKDQSDDDLKDLSLKRNQTVVDVAVLTSPLODEKYESYV 2261

RESULT 10
AAB31367
ID AAB31367 standard; Protein; 2261 AA.
XX
AC AAB31367;
XX
DT 20-Apr-2001 (first entry)
XX
DE Amino acid sequence of ABC1 polypeptide from Tangier disease patient.
XX
KW Human: adenosine triphosphate binding cassette protein 1; ABC1;
KW apolipoprotein-mediated mobilisation; cholesterol; Tangier disease;
KW chromosome 9q22-9q31; heart disease; hypercholesterolemia;
KW atherosclerosis; cholesterol transport.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 587
FT PT /note="this is changed from Arg to Trp in Tangier disease"
XX
PN NC0200078971-A2.
XX
PD 28-DEC-2000.
XX
PF 16-JUN-2000; 2000MO-US16591.
XX
PR 18-JUN-1999; 9905-0140264.
PR 14-SEP-1999; 9905-0153872.
PR 19-NOV-1999; 9905-0166573.
XX
PA (CVTH-) CV THERAPEUTICS INC.
PA (UNITW) UNIT WASHINGTON.
PI Lawn RM, Wade D, Oram JF, Garvin M;
XX
WP1: 2001-137811/14.
DR N-PSDB: AAF24708.
XX
PT Adenosine triphosphate (ATP) binding cassette protein (ABC) 1
PT polynucleotides and polypeptides, useful for treatment of heart disease
PT and other disorders associated with hypercholesterolemia and
PT atherosclerosis -

```


XX PS Claim 28; Page 172-187; 211pp; English.
 XX CC The present sequence represents a human adenosine triphosphate (ATP)
 CC binding cassette protein (ABC) 1 polypeptide, and is isolated from
 CC a Tangier disease patient. ABC1 resides in cell membranes and utilises
 CC ATP hydrolysis to transport a wide variety of substrates across the
 CC plasma membrane. ABC1 is a pivotal protein in the apolipoprotein-mediated
 CC mobilisation of intracellular cholesterol stores. ABC1 is defective in
 CC Tangier disease, a genetic disorder characterised by abnormal
 CC HDL-cholesterol metabolism. The ABC1 gene is localised to chromosome
 CC 9q22-9q31. The ABC1 genes and proteins are useful for developing
 CC pharmaceutical agents for the treatment of heart disease and other
 CC disorders associated with hypercholesterolemia and atherosclerosis. The
 CC genes are useful for developing screening assays to screen for compounds
 CC that regulate the expression of genes associated with cholesterol
 CC transport. The genes and proteins are also useful for are also useful
 CC as diagnostic indicators of cardiovascular disease and other disorders
 CC associated with hypercholesterolemia.
 XX CC
 SQ Sequence 2261 AA;
 Query Match 99.9%; Score 11789; DB 22; Length 2261;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2260; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 721 FLVFAVAVTIIQCFIISTLFESRANLAAACGGIYFTLYPLVLCVAMQDVGFTLKIFAS 780
 DB 721 FLVFAVAVTIIQCFIISTLFESRANLAAACGGIYFTLYPLVLCVAMQDVGFTLKIFAS 780
 QY 781 LLSVAGFCGEYALFEEOGICVQWQNLFESEYEEGFMULTISIMMLFDFTLYGVMTW 840
 DB 781 LLSVAGFCGEYALFEEOGICVQWQNLFESEYEEGFMULTISIMMLFDFTLYGVMTW 840
 QY 841 YIEAVFGQVGIPIPPWFFPCPKSYWFGEEDEKSHPSNOKRMSCEICMEEPHLKIGVS 900
 DB 841 YIEAVFGQVGIPIPPWFFPCPKSYWFGEEDEKSHPSNOKRMSCEICMEEPHLKIGVS 900
 QY 901 IQLNVKYYRDMKAVADGLALNFYEQIITSLFNGAGKTTWISILTGLFPPTSGAVYII 960
 DB 901 IQLNVKYYRDMKAVADGLALNFYEQIITSLFNGAGKTTWISILTGLFPPTSGAVYII 960
 QY 961 GKDIRSEMTIRQNLGYCPQHNVLFDMLYVEEHTWFARLKGLSEKHVAKEMQMLDVC 1020
 DB 961 GKDIRSEMTIRQNLGYCPQHNVLFDMLYVEEHTWFARLKGLSEKHVAKEMQMLDVC 1020
 QY 1021 LPSSKLKSKTSQLSGQMQRKLSVALAFVGSKVYIIDEPTAGVDPYSRRGIWELLKTYRQ 1080
 DB 1021 LPSSKLKSKTSQLSGQMQRKLSVALAFVGSKVYIIDEPTAGVDPYSRRGIWELLKTYRQ 1080
 QY 1081 GRTIILSTHMDADVIGDRIAIISHGKLCVGSGLFLAKQDGTGYLLVKKDVESSLS 1140
 DB 1081 GRTIILSTHMDADVIGDRIAIISHGKLCVGSGLFLAKQDGTGYLLVKKDVESSLS 1140
 QY 1141 SCRNSSTFVSLKKEDESVSSGSDAGSGDHESTLTLDVSALSNIIRKVSARLVEDI 1200
 DB 1141 SCRNSSTFVSLKKEDESVSSGSDAGSGDHESTLTLDVSALSNIIRKVSARLVEDI 1200
 QY 1201 GHELTLYVLPYBAKEGAFVELFHEIDRLSDLGISYISSETTLEETFLKVAEESGVDAE 1260
 DB 1201 GHELTLYVLPYBAKEGAFVELFHEIDRLSDLGISYISSETTLEETFLKVAEESGVDAE 1260
 QY 1261 TSDGTLPARRRRRAFGDKOSCLRPFTEDDAADPDSDIDESBETDLSGMDKGSYOK 1320
 DB 1261 TSDGTLPARRRRRAFGDKOSCLRPFTEDDAADPDSDIDESBETDLSGMDKGSYOK 1320
 QY 1321 GWKLTQOQFVALMKRLLIARRSKRGFFAOIVLPAVVCIALVFSILVPPFGKYPSLEIQ 1380
 DB 1321 GWKLTQOQFVALMKRLLIARRSKRGFFAOIVLPAVVCIALVFSILVPPFGKYPSLEIQ 1380
 QY 1381 PMATNEOYTFVSNAPEDDTGTELLNALTKDPGFGTRCMGNGIIPDTPOAGEEWTTPAR 1440
 DB 1381 PMATNEOYTFVSNAPEDDTGTELLNALTKDPGFGTRCMGNGIIPDTPOAGEEWTTPAR 1440
 QY 1441 VPQITMDLFQNGNWTMGNPSPACCCSSDKIKKMLPVCPPAGGIPPRQKONTADILQDL 1500
 DB 1441 VPQITMDLFQNGNWTMGNPSPACCCSSDKIKKMLPVCPPAGGIPPRQKONTADILQDL 1500
 QY 1501 TGRNISDYLVKTYVQIIIAKSLKINKIWNFEFRYGGFSLGVSNTQALPPSOEVDNAIKOMK 1560
 DB 1501 TGRNISDYLVKTYVQIIIAKSLKINKIWNFEFRYGGFSLGVSNTQALPPSOEVDNAIKOMK 1560
 QY 1561 HLKIAKSSADRFNLNSLGRMTGIDTFRNNKVKVFPNNKGHAISSFLNVIINATILRANLQK 1620
 DB 1561 HLKIAKSSADRFNLNSLGRMTGIDTFRNNKVKVFPNNKGHAISSFLNVIINATILRANLQK 1620
 QY 1621 GENPSHGITAIFNHPNLUTRQOOLSEVALMTTSVDVLVSIQVIFAMSFVPSFVAFILOER 1680
 DB 1621 GENPSHGITAIFNHPNLUTRQOOLSEVALMTTSVDVLVSIQVIFAMSFVPSFVAFILOER 1680
 QY 1681 VSKAKHLQFISGVKPVLYLWLSNFWDMCNVVPATLVIIIFIOFOOKSYVSSTNLEPVIAL 1740
 DB 1681 VSKAKHLQFISGVKPVLYLWLSNFWDMCNVVPATLVIIIFIOFOOKSYVSSTNLEPVIAL 1740
 QY 1741 LLLLYGSIPIPMWPAFVKRIPSTAVVLTSVNLFGINGSVATVLEFTDNKLNIN 1800
 DB 1741 LLLLYGSIPIPMWPAFVKRIPSTAVVLTSVNLFGINGSVATVLEFTDNKLNIN 1800


```

OY 1801 DILKSVFLIFPHFCGLRGLIDMVKNOAMADALERFGENREVSPLSMIDVGRNLFAMAVEG 1860
DB 1801 DILKSVFLIFPHFCGLRGLIDMVKNOAMADALERFGENREVSPLSMIDVGRNLFAMAVEG 1860
OY 1861 VFFLITVLIQYREFIRPRPVNAKISPLINDEDEDVRRERQRIIDGGGONDILEIKELTKI 1920
DB 1861 VFFLITVLIQYREFIRPRPVNAKISPLINDEDEDVRRERQRIIDGGGONDILEIKELTKI 1920
OY 1921 YRRKKRPVADRICVGIIPGECFGLLVGAGKSSSFKNLTJDTVTTRCDATLNKNSILSN 1960
DB 1921 YRRKKRPVADRICVGIIPGECFGLLVGAGKSSSFKNLTJDTVTTRCDATLNKNSILSN 1960
OY 1981 IHEVQNNNGCYCPQFPAITELLTGRHVEFFALLRGVPEKEVGKYGEMAIRKLGIVKYEK 2040
DB 1981 IHEVQNNNGCYCPQFPAITELLTGRHVEFFALLRGVPEKEVGKYGEMAIRKLGIVKYEK 2040
OY 2041 YAGNYSGGNKKRLSTAMALIGGPVVFIDEPTGMDPKARRFLMNCALSVYKEGRSVLT 2100
DB 2041 YAGNYSGGNKKRLSTAMALIGGPVVFIDEPTGMDPKARRFLMNCALSVYKEGRSVLT 2100
OY 2101 SHSMECEALCTRMAIMVNGRRCISGVQHLKNRFGDGYTTIVRAGSNPDLKPYODEFG 2160
DB 2101 SHSMECEALCTRMAIMVNGRRCISGVQHLKNRFGDGYTTIVRAGSNPDLKPYODEFG 2160
OY 2161 LAFPGSVLKEKRRNMLQYQLPSSLSLARIFSLSQSKRLHIEDYSVSQTTLDQVFVNF 2220
DB 2161 LAFPGSVLKEKRRNMLQYQLPSSLSLARIFSLSQSKRLHIEDYSVSQTTLDQVFVNF 2220
OY 2221 AKDQSDHKLKDLSTLHKNTVVDAVLTSLFLODEKVESYV 2261
DB 2221 AKDQSDHKLKDLSTLHKNTVVDAVLTSLFLODEKVESYV 2261

RESULT 11
ID AAB38111 standard; Protein: 2261 AA.
XX
XX AAB38111;
AC
XX
XX 29-JAN-2001 (first entry)
DE
XX Human ABC1 cholesterol transporter mutant, V771M.
XX
XX Human ABC1 cholesterol transporter; chromosome 9q31;
XX ATP-binding cassette; HDL deficiency disorder; high density lipoprotein;
XX Tangier disease; TD; familial HDL deficiency; FHA; polymorphism;
XX cardiovascular disease; coronary artery disease; coronary restenosis;
XX cerebrovascular disease; peripheral vascular disease;
XX Alzheimer's disease; Niemann-Pick disease; Huntington's disease;
XX X-linked adrenoleukodystrophy; cancer; gene therapy; genetic diagnosis;
XX prognosis; prophylaxis; drug screening; transgenic animal; mutant;
XX
XX Homo sapiens.
OS
XX
XX WO200055318-A2.
XX
XX 21-SEP-2000.
XX
XX 15-MAR-2000; 2000WO-1B00532.
XX
XX 15-MAR-1999; 9905-0124702.
XX 08-JUN-1999; 9905-0138048.
XX 17-JUN-1999; 9905-0139600.
XX 01-SEP-1999; 9905-0151977.
XX
XX (UYBR-) UNIT BRITISH COLUMBIA.
XX (XENO-) XENON BIORESEARCH INC.
XX
XX Hayden MR, Wilson AR, Pimstone SN;
XX
XX WPI; 2000-587528/55.
XX

```

```

PT New ABC1 polypeptide is useful for treating diseases associated with
PT ABC1 biological activity, e.g. Alzheimer's disease, Huntington's
PT disease and cancer -
XX
XX Examples: Page -: 229pp; English.
XX
XX The invention relates to the human ABC1 cholesterol transporter protein
XX (B38082) and to nucleic acid sequences (c69120) which encode it. ABC1 is
XX a member of the ATP-binding cassette (ABC transporter) superfamily of
XX proteins, and plays a crucial role in cholesterol transport, particularly
XX intracellular cholesterol trafficking in monocytes and fibroblasts, being
XX involved in cholesterol efflux from the cell. The gene encoding ABC1 is
XX located on chromosome 9q31, and mutations in this gene are associated
XX with two genetic HDL (high density lipoprotein) deficiency disorders,
XX Tangier disease (TD) and familial HDL deficiency (FHA). These diseases
XX are distinguishable in that TD is an autosomal recessive disorder, while
XX FHA is inherited as an autosomal dominant trait. Low levels of HDL ("good
XX cholesterol") in the blood correlate with a high risk of cardiovascular
XX disease, particularly coronary artery disease, but also cerebrovascular
XX disease, coronary restenosis, and peripheral vascular disease.
XX Conversely, a high level of HDL has protective effects against
XX cardiovascular disease. The invention provides genetic constructs and
XX transgenic cells and non-human animals comprising human ABC1 nucleic
XX acids, and methods of gene therapy for the treatment or prevention of
XX cardiovascular disease comprising the administration of an expression
XX vector encoding ABC1 or an active fragment thereof. The invention also
XX encompasses compounds which mimic ABC1 activity, compounds which
XX stimulate ABC1 expression and methods of screening for such compounds.
XX It further relates to methods for determining whether for a patient has an
XX increased risk for cardiovascular disease due to polymorphisms in the
XX ABC1 gene. Human ABC1 proteins and nucleotides can be used to treat
XX or prevent cardiovascular disease, especially coronary artery disease,
XX cerebrovascular disease, coronary restenosis or peripheral vascular
XX disease. They may also be used in the treatment of diseases associated
XX with ABC1 biological activity, such as Alzheimer's disease, Niemann-Pick
XX disease, Huntington's disease, X-linked adrenoleukodystrophy and cancer.
XX The invention specifically excludes proteins with the exact amino acid
XX sequences of GenBank Accession No: CAA10005.1 and X75926, and the nucleic
XX acid with the exact sequence as GenBank Accession No: AJ012376.1. The
XX present sequence represents a mutant human ABC1 cholesterol transporter
XX associated with an altered cholesterol level and therefore an altered
XX risk of cardiovascular disease.
XX Note: The present sequence is not shown in the specification, but is
XX derived from the native human ABC1 shown on pages 152-157.
XX
XX Sequence 2261 AA:
XX
XX Query Match 99.9%; Score 11786; DB 21; Length 2261;
XX Best Local Similarity 99.8%; Pred. No. 0;
XX Matches 2257; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
OY 1 MACWPQLRLILMKNLTFRRQTCOLLLEVAMPFLTIFILISVRLSPYRDOHECHFPNKA 60
DB 1 MACWPQLRLILMKNLTFRRQTCOLLLEVAMPFLTIFILISVRLSPYRDOHECHFPNKA 60
OY 61 MPSAGTLPWVGIIICNANNPCFRYPTEGAPGVGVGNFNKSIVARLFSFARLLLYSQKDT 120
DB 61 MPSAGTLPWVGIIICNANNPCFRYPTEGAPGVGVGNFNKSIVARLFSFARLLLYSQKDT 120
OY 121 SMKDMRKVRLTLQOIKKSSNLKIDFLVNDPEFSGFLYHNLSLPKSTGVKMLRADYILH 180
DB 121 SMKDMRKVRLTLQOIKKSSNLKIDFLVNDPEFSGFLYHNLSLPKSTGVKMLRADYILH 180
OY 181 KYFLQGYOLHLTSLCNGSKSEMIQIDQEVSELGCPREKLAAREVLRSNMDILKPIIL 240
DB 181 KYFLQGYOLHLTSLCNGSKSEMIQIDQEVSELGCPREKLAAREVLRSNMDILKPIIL 240
OY 241 RLINSTSPFPEKLEAETKTLHLISGLTLAGLFSMRKSDMRQEVMLTNNVSSSSSTQI 300
DB 241 RLINSTSPFPEKLEAETKTLHLISGLTLAGLFSMRKSDMRQEVMLTNNVSSSSSTQI 300
OY 301 YQAVSRIVCGHPEGGLIKLSLNMVEDNNYKALFGNGTEEDAEFFYDNGSTTPPCNDLMLK 360
DB 301 YQAVSRIVCGHPEGGLIKLSLNMVEDNNYKALFGNGTEEDAEFFYDNGSTTPPCNDLMLK 360

```

Db 301 YQAVSRIVCGHPEGGLIKSLNWEEDNNKALFEGNGCTEEDAEFTFYDNSTTPYCNDIMK 360
 QY 361 NLESSPLSRITWKALKPLLVGRILVTPDT PATROYMAEVNKTFOELAVFHLDEGMWELLS 420
 Db 361 NLESSPLSRITWKALKPLLVGRILVTPDT PATROYMAEVNKTFOELAVFHLDEGMWELLS 420
 QY 421 PKIWMFEMSOEMDLVNRMLDSRDHFWEOQLDMDTAD IYAF LAKHEDVOSSNGS 480
 Db 421 PKIWMFEMSOEMDLVNRMLDSRDHFWEOQLDMDTAD IYAF LAKHEDVOSSNGS 480
 QY 481 VYTWBAEFNETNOATRTISREMECVNLKLEPIATEVWLJNKSMELLDERKFMAGIYFTG 540
 Db 481 VYTWBAEFNETNOATRTISREMECVNLKLEPIATEVWLJNKSMELLDERKFMAGIYFTG 540
 QY 541 ITPGSELPHVHKYKTRMDIDNVEFTNKIKDGYWDPGRADPFEDMRYVWGFAVLODGV 600
 Db 541 ITPGSELPHVHKYKTRMDIDNVEFTNKIKDGYWDPGRADPFEDMRYVWGFAVLODGV 600
 QY 601 FOAIIIRVLGTGTEKTVGMQMPYPCYVDJIEFRVMSRSMLEMTLAMITYSVANIYKGIY 660
 Db 601 FOAIIIRVLGTGTEKTVGMQMPYPCYVDJIEFRVMSRSMLEMTLAMITYSVANIYKGIY 660
 QY 661 YEKEARLKEKTRIMGLDINSILMFSWFTSSILPLVSAGLVILKGNLIPYSDPSVFEV 720
 Db 661 YEKEARLKEKTRIMGLDINSILMFSWFTSSILPLVSAGLVILKGNLIPYSDPSVFEV 720
 QY 721 FLVSAVAVTIIIOCLIFLSTRANLAAACGGIYFTLYLPVYLCVAMODYVFTLKIFAS 780
 Db 721 FLVSAVAVTIIIOCLIFLSTRANLAAACGGIYFTLYLPVYLCVAMODYVFTLKIFAS 780
 QY 781 LLSPAFGGCEYFALFEQSIGVOMDNLFPSPVEEDGFNTTISIMLFPOTELIXGMW 840
 Db 781 LLSPAFGGCEYFALFEQSIGVOMDNLFPSPVEEDGFNTTISIMLFPOTELIXGMW 840
 QY 841 YIEAVFPGQYCIIPRPWYFPCYKSYWGESEDEKSHPSNOKRMEICIMEEPTHLKIGVS 900
 Db 841 YIEAVFPGQYCIIPRPWYFPCYKSYWGESEDEKSHPSNOKRMEICIMEEPTHLKIGVS 900
 QY 901 IONLVKVRDGMKVAVDGLALNFYEGOTSPFGHNGAKTWTMSILTLGLFPPTSGTAYIL 960
 Db 901 IONLVKVRDGMKVAVDGLALNFYEGOTSPFGHNGAKTWTMSILTLGLFPPTSGTAYIL 960
 QY 961 GKDIRSEKSTIRONLGVCPHNVLFMDLTVEEHIMFYARLKGLSEKHVAKEMEMALDVG 1020
 Db 961 GKDIRSEKSTIRONLGVCPHNVLFMDLTVEEHIMFYARLKGLSEKHVAKEMEMALDVG 1020
 QY 1021 LPSSKLKSTQSLSGGMQRKLSVALAVGSKVYITDEPTAGVDPYSRRGIWELLKRYO 1080
 Db 1021 LPSSKLKSTQSLSGGMQRKLSVALAVGSKVYITDEPTAGVDPYSRRGIWELLKRYO 1080
 QY 1081 GRITITLSTHMHDEADVLCGRITAITSHGKLCVGSFLKNOLOGVYLTLYKKDVESLS 1140
 Db 1081 GRITITLSTHMHDEADVLCGRITAITSHGKLCVGSFLKNOLOGVYLTLYKKDVESLS 1140
 QY 1141 SCRNSSTVSTYLKEDSVSSSSDAGLSDHESDTLTIDVSAISNLRKHVSARLVEDI 1200
 Db 1141 SCRNSSTVSTYLKEDSVSSSSDAGLSDHESDTLTIDVSAISNLRKHVSARLVEDI 1200
 QY 1201 GHELTJYLPYEAKEGAFVFLFHEIDRLSDLGISGISETTLLEFLKVAEESGVDAE 1260
 Db 1201 GHELTJYLPYEAKEGAFVFLFHEIDRLSDLGISGISETTLLEFLKVAEESGVDAE 1260
 QY 1261 TSDGTLPARNRRAFGKOSGLRPFTEDDAADPNDSIDPESRETDLSGMDKGSYQAK 1320
 Db 1261 TSDGTLPARNRRAFGKOSGLRPFTEDDAADPNDSIDPESRETDLSGMDKGSYQAK 1320
 QY 1321 GWKLTQOQFVALLMKRLLIARRSKGFFAOIYLPVAVFCIALVFSILVPPGKYPSLEIQ 1380
 Db 1321 GWKLTQOQFVALLMKRLLIARRSKGFFAOIYLPVAVFCIALVFSILVPPGKYPSLEIQ 1380
 QY 1381 PMWYNEQYTVSNDAPDPTGTELLNALT KDPEGRCHMGKNIPPTPOAGGEEMWTAP 1440
 Db 1381 PMWYNEQYTVSNDAPDPTGTELLNALT KDPEGRCHMGKNIPPTPOAGGEEMWTAP 1440

QY 1441 VPOTIMDLFONGWMTQNPSPACOCSSDKIKMLPVCPPGAGGLPPPOKONTADITLODL 1500
 Db 1441 VPOTIMDLFONGWMTQNPSPACOCSSDKIKMLPVCPPGAGGLPPPOKONTADITLODL 1500
 QY 1501 TGRNISDYLVKTYVOJIIANSKANKIVNEFRYGFSLGVSNTOALPPSOEVNDAIKQMK 1560
 Db 1501 TGRNISDYLVKTYVOJIIANSKANKIVNEFRYGFSLGVSNTOALPPSOEVNDAIKQMK 1560
 QY 1561 HLKLAODSSADRFNLNSIGRMGTLOTRNNVKNVFNKKGHAISSPLNTVNNAILRANIOK 1620
 Db 1561 HLKLAODSSADRFNLNSIGRMGTLOTRNNVKNVFNKKGHAISSPLNTVNNAILRANIOK 1620
 QY 1621 GENPSHYGITAFAFNHPLNLTQOOLSEVALMTSYDVLSICVIFAMSFVAPASVVEILOER 1680
 Db 1621 GENPSHYGITAFAFNHPLNLTQOOLSEVALMTSYDVLSICVIFAMSFVAPASVVEILOER 1680
 QY 1681 VSKAKHLQFISGVKPYIYIWLNSFWDMCNVVPATLVIIIFICFOOKSVYSTNLPVAL 1740
 Db 1681 VSKAKHLQFISGVKPYIYIWLNSFWDMCNVVPATLVIIIFICFOOKSVYSTNLPVAL 1740
 QY 1741 LLLLYGMSITPLMYPASFEFKIPSTAYVVLTSVNLFIGINSVATFVLELFTDNKLNNIN 1800
 Db 1741 LLLLYGMSITPLMYPASFEFKIPSTAYVVLTSVNLFIGINSVATFVLELFTDNKLNNIN 1800
 QY 1801 DILKSVFLIFPHECLGRGLDMVKNQAMADALERGENRFSPLSMDLVGRMLFMAVEG 1860
 Db 1801 DILKSVFLIFPHECLGRGLDMVKNQAMADALERGENRFSPLSMDLVGRMLFMAVEG 1860
 QY 1861 VVFFELTYVLIQYRFILRPBPVNAKLSPLNDEDEDVRRERORILDDGGONDILEIKELTKI 1920
 Db 1861 VVFFELTYVLIQYRFILRPBPVNAKLSPLNDEDEDVRRERORILDDGGONDILEIKELTKI 1920
 QY 1921 YRRKRKPAVDRLICVGLPPECFGLGVNAGKSTFFKMLTGTPTYTRGDAPLNKSTILSN 1980
 Db 1921 YRRKRKPAVDRLICVGLPPECFGLGVNAGKSTFFKMLTGTPTYTRGDAPLNKSTILSN 1980
 QY 1981 IHEVHONMGYCPQFDATIELLTGREHVEFFALLRGVPEKEVGKVGEMAIRKLGIVKGEK 2040
 Db 1981 IHEVHONMGYCPQFDATIELLTGREHVEFFALLRGVPEKEVGKVGEMAIRKLGIVKGEK 2040
 QY 2041 YAGNVSCKNRKRLSTAMALIGBPVVFIDEPTTGMDPKARRIIMCALSVYKEGSSVLT 2100
 Db 2041 YAGNVSCKNRKRLSTAMALIGBPVVFIDEPTTGMDPKARRIIMCALSVYKEGSSVLT 2100
 QY 2101 SHSMECEALCTRMAIMVNGFRCLGSVOHLKNRFGDGYTIVIRIAGSNPDLKPVQDFPG 2160
 Db 2101 SHSMECEALCTRMAIMVNGFRCLGSVOHLKNRFGDGYTIVIRIAGSNPDLKPVQDFPG 2160
 QY 2161 LAFPGSVILKEKRRNMLQYQLPSSLSLARIFSLISQSKRRLHIEDYVSQTTLDQVEVNF 2220
 Db 2161 LAFPGSVILKEKRRNMLQYQLPSSLSLARIFSLISQSKRRLHIEDYVSQTTLDQVEVNF 2220
 QY 2221 AKDQSDDDHLKDLSLHKNOTYVDVAVLTSFLQDEKVKSSYV 2261
 Db 2221 AKDQSDDDHLKDLSLHKNOTYVDVAVLTSFLQDEKVKSSYV 2261
 RESULT 12
 AAB38114
 ID AAB38114 standard; Protein: 2261 AA.
 XX
 AC AAB38114;
 XX
 DT 29-JAN-2001 (first entry)
 XX
 DE Human ABC1 cholesterol transporter mutant, E1172D.
 XX
 Human ABC1 cholesterol transporter; chromosome 9q31;
 KW ATP-binding cassette; HDL deficiency disorder; high density lipoprotein;
 KW Tangier disease; Td; familial HDL deficiency; FNA; polymorphism;
 KW cardiovascular disease; coronary artery disease; coronary restenosis;
 KW cerebrovascular disease; peripheral vascular disease;

KM Alzheimer's disease; Niemann-Pick disease; Huntington's disease;
 KM X-linked adrenoleukodystrophy; cancer; gene therapy; genetic diagnosis;
 KM proinosis; prophylaxis; drug screening; transgenic animal; mutant;
 KM mutant.
 OS Homo sapiens.
 XX MO20005318-A2.
 XX PD 21-SEP-2000.
 XX 15-MAR-2000; 2000WO-1B00532.
 XX PF 15-MAR-1999; 99US-0124702.
 XX PR 08-JUN-1999; 99US-0138048.
 XX PR 17-JUN-1999; 99US-0139600.
 XX PR 01-SEP-1999; 99US-0151977.
 XX (UYBR-) UNIV BRITISH COLUMBIA.
 PA (XENO-) XENON BIORESEARCH INC.
 XX Hayden MR, Wilson AR, Pimstone SN;
 PI WPI; 2000-587528/55.
 DR New ABC1 polypeptide is useful for treating diseases associated with
 PT ABC1 biological activity, e.g. Alzheimer's disease, Huntington's
 PT disease and cancer.
 PS Examples; Page -: 229pp; English.
 XX The invention relates to the human ABC1 cholesterol transporter protein
 CC (B36082) and to nucleic acid sequences (C69120) which encode it. ABC1 is
 CC a member of the ATP-binding cassette (ABC transporter) superfamily of
 CC proteins, and plays a crucial role in cholesterol transport, particularly
 CC intracellular cholesterol trafficking in monocytes and fibroblasts, being
 CC involved in cholesterol efflux from the cell. The gene encoding ABC1 is
 CC located on chromosome 9q31, and mutations in this gene are associated
 CC with two genetic HDL (high density lipoprotein) deficiency disorders,
 CC Tangier disease (TD) and familial HDL deficiency (FHA). These diseases
 CC are distinguishable in that TD is an autosomal recessive disorder, while
 CC FHA is inherited as an autosomal dominant trait. Low levels of HDL ("good
 CC cholesterol") in the blood correlate with a high risk of cardiovascular
 CC disease, particularly coronary artery disease, but also cerebrovascular
 CC disease, coronary stenosis, and peripheral vascular disease.
 CC Conversely, a high level of HDL has protective effects against
 CC cardiovascular disease. The invention provides genetic constructs and
 CC transgenic cells and non-human animals comprising human ABC1 nucleic
 CC acids, and methods of gene therapy for the treatment or prevention of
 CC cardiovascular disease comprising the administration of an expression
 CC vector encoding ABC1 or an active fragment thereof. The invention also
 CC encompasses compounds which mimic ABC1 activity, compounds which
 CC stimulate ABC1 expression and methods of screening for such compounds.
 CC It further relates to methods for determining whether a patient has an
 CC increased risk for cardiovascular disease due to polymorphisms in the
 CC ABC1 gene. Human ABC1 proteins and nucleotides can be used to treat
 CC or prevent cardiovascular disease, especially coronary artery disease,
 CC cerebrovascular disease, coronary stenosis or peripheral vascular
 CC disease. They may also be used in the treatment of diseases associated
 CC with ABC1 biological activity, such as Alzheimer's disease, Niemann-Pick
 CC disease, Huntington's disease, X-linked adrenoleukodystrophy and cancer.
 CC The invention specifically excludes proteins with the exact amino acid
 CC sequences of Genbank Accession No: CAA10005.1 and X75926, and the nucleic
 CC acid with the exact sequence as Genbank Accession No: AJ012376.1. The
 CC present sequence represents a mutant human ABC1 cholesterol transporter
 CC associated with an altered cholesterol level and therefore an altered
 CC risk of cardiovascular disease.
 CC Note: The present sequence is not shown in the specification, but is
 CC derived from the native human ABC1 shown on pages 152-157.
 XX
 XX Sequence 2261 AA:
 SQ Query Match 99.9%; Score 11786; DB 21; Length 2261;

Best Local Similarity 99.8%; Pred. No. 0;
 Matches 2257; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MACHPQRLRLMKLNFRRQTCQLLEVAWPLFIFILISVLSLSPPEQHECHPNKA 60
 Db 1 MACHPQRLRLMKLNFRRQTCQLLEVAWPLFIFILISVLSLSPPEQHECHPNKA 60
 QY 61 MPSAGTLPWVGIIICNANNPCFRPPGEPAGVGVGNFNKSTVARLSDARLLYSQKPT 120
 Db 61 MPSAGTLPWVGIIICNANNPCFRPPGEPAGVGVGNFNKSTVARLSDARLLYSQKPT 120
 QY 121 SMKDMRKVLRTLOQIKKSSNKLQDFLVNMFESGFLYHNLSIPKSTYDKMLRADVILH 180
 Db 121 SMKDMRKVLRTLOQIKKSSNKLQDFLVNMFESGFLYHNLSIPKSTYDKMLRADVILH 180
 QY 181 KVFLOGYQLHLTSLCNSKSEDEMIQLQDOYSELGCPKPKKLAARVLRSMNDILKPLT 240
 Db 181 KVFLOGYQLHLTSLCNSKSEDEMIQLQDOYSELGCPKPKKLAARVLRSMNDILKPLT 240
 QY 241 RTLNSTSPPSKELAEATKTLHSLGTLAQLFMSRMSDMROEVLTVNVSSSSSTQI 300
 Db 241 RTLNSTSPPSKELAEATKTLHSLGTLAQLFMSRMSDMROEVLTVNVSSSSSTQI 300
 QY 301 YQAVSRIVCGHPEGGLKTRSLNMYEDNNYKALFGNGTEDEAETFDNSTTPYCNDLAK 360
 Db 301 YQAVSRIVCGHPEGGLKTRSLNMYEDNNYKALFGNGTEDEAETFDNSTTPYCNDLAK 360
 QY 361 NLESSPLSRILWKALKPLVGLKILYTPDTPATROVAAEVKTKTQELAVFHDLGMMEEIS 420
 Db 361 NLESSPLSRILWKALKPLVGLKILYTPDTPATROVAAEVKTKTQELAVFHDLGMMEEIS 420
 QY 421 PKITWPMENSOEMLVMLLDSRDHFMEOQLDGLMTWODIVAFIAKHPEDVOSSNGS 480
 Db 421 PKITWPMENSOEMLVMLLDSRDHFMEOQLDGLMTWODIVAFIAKHPEDVOSSNGS 480
 QY 481 VYTRAEAFNEINQAIRTISREMECVNLKLEPATEVWMLINKSMELDERKFMAGIVFTG 540
 Db 481 VYTRAEAFNEINQAIRTISREMECVNLKLEPATEVWMLINKSMELDERKFMAGIVFTG 540
 QY 541 ITPPSIELPHHVKKIKIMDDINVERTKIKDGYNDPPRADPPEDMKYVWGGAIVLQDVY 600
 Db 541 ITPPSIELPHHVKKIKIMDDINVERTKIKDGYNDPPRADPPEDMKYVWGGAIVLQDVY 600
 QY 601 EQAIRVLTGTEKTKGYVMQMPYCYVDIIFLRVMSRSPLEMTLAMIYSVAVIIKGIY 660
 Db 601 EQAIRVLTGTEKTKGYVMQMPYCYVDIIFLRVMSRSPLEMTLAMIYSVAVIIKGIY 660
 QY 661 YEKEARKLETMRIMGDLSILMFSWFTSSLIPLLVSAGLLVIIKGNLLPYSDPSVVFV 720
 Db 661 YEKEARKLETMRIMGDLSILMFSWFTSSLIPLLVSAGLLVIIKGNLLPYSDPSVVFV 720
 QY 721 FLSVFAVVTIIQCFILSTIFSRANLAAACGIIYFTLYLPVLCVAMQDVVGFTLKIFAS 780
 Db 721 FLSVFAVVTIIQCFILSTIFSRANLAAACGIIYFTLYLPVLCVAMQDVVGFTLKIFAS 780
 QY 781 LLSPAVFGCEYFALPEEOGIGVOMDNLFPESPYEEDGFNLTTISIMMLDFTFLYGMTV 840
 Db 781 LLSPAVFGCEYFALPEEOGIGVOMDNLFPESPYEEDGFNLTTISIMMLDFTFLYGMTV 840
 QY 841 YIEAVFPQGYIIPRWYFPCTKSYWFGESDEKSHPSGNOKRMSEICMEEBPHILKLGVS 900
 Db 841 YIEAVFPQGYIIPRWYFPCTKSYWFGESDEKSHPSGNOKRMSEICMEEBPHILKLGVS 900
 QY 901 IONLVKVRDGMKVAVDLALNFEQGTTSFLGHNGAKTTTMSILNGLPPTSGTAYIL 960
 Db 901 IONLVKVRDGMKVAVDLALNFEQGTTSFLGHNGAKTTTMSILNGLPPTSGTAYIL 960
 QY 961 GKDIRSEMTIRONGVCPQHNVLFDMLTVEHTMFAARLKGISEKKHYKAEQEMALDVG 1020
 Db 961 GKDIRSEMTIRONGVCPQHNVLFDMLTVEHTMFAARLKGISEKKHYKAEQEMALDVG 1020
 QY 1021 LPSSKLKSKTSQSLSGMOKRLSVALAFVGSKVVLDEPTAGVDPYSRGIWELLKRYK 1080
 Db 1021 LPSSKLKSKTSQSLSGMOKRLSVALAFVGSKVVLDEPTAGVDPYSRGIWELLKRYK 1080

```

Db 1021 LPSKLSKTSQLSGGMQRKLSVALFVGSKVYLDEPTAGVDPRSRGRIWEILLKTRQ 1080
QY 1081 GRTIILSTHHMEADVYGRIRIISHGKLCYCGSSLEFLKNOLGTYYTLTVKKDVSSLS 1140
Db 1081 GRTIILSTHHMEADVYGRIRIISHGKLCYCGSSLEFLKNOLGTYYTLTVKKDVSSLS 1140
QY 1141 SCRNSSTSYLAKKEDSVSSQSSDAGLSDHSDTLTIDVSAISNLTKRHVSEALVEDI 1200
Db 1141 SCRNSSTSYLAKKEDSVSSQSSDAGLSDHSDTLTIDVSAISNLTKRHVSEALVEDI 1200
QY 1201 GHELYVLYPEAKGAEVLEFHEIDRLSDLGISSYGISETTLEIFLKAEEGVDAE 1260
Db 1201 GHELYVLYPEAKGAEVLEFHEIDRLSDLGISSYGISETTLEIFLKAEEGVDAE 1260
QY 1261 TSDGTLPAARNRRAFGDQSCIRPTEDDADPNDSIDIPESREDDLSDGDKGSYQV 1320
Db 1261 TSDGTLPAARNRRAFGDQSCIRPTEDDADPNDSIDIPESREDDLSDGDKGSYQV 1320
QY 1321 GMLTQOQFVALLMKRLILARSRKGFPAQIYLPVAFVCIALVFSLIYPPGKYPSELO 1380
Db 1321 GMLTQOQFVALLMKRLILARSRKGFPAQIYLPVAFVCIALVFSLIYPPGKYPSELO 1380
QY 1381 PMWNEOYTFFVSNDAPEDTGILELINALTKDPGEFGRMEGNPIPDTPCQAGEEWTAP 1440
Db 1381 PMWNEOYTFFVSNDAPEDTGILELINALTKDPGEFGRMEGNPIPDTPCQAGEEWTAP 1440
QY 1441 VPQITMDLFFQNGNMTQNPSPACQSSDKIKMLPYCPGAGGPPPRKONTADIIODL 1500
Db 1441 VPQITMDLFFQNGNMTQNPSPACQSSDKIKMLPYCPGAGGPPPRKONTADIIODL 1500
QY 1501 TGRNISDYLVKTYVOIIAKSLKNKIWNEFRYGGSLGVSNTQALPQSEVNDARIKQMK 1560
Db 1501 TGRNISDYLVKTYVOIIAKSLKNKIWNEFRYGGSLGVSNTQALPQSEVNDARIKQMK 1560
QY 1561 HKLAKDSSADDFLNSLGRFMTGIDTRNNVYVWFMNKKMHAISFLVNTANAILRANLOK 1620
Db 1561 HKLAKDSSADDFLNSLGRFMTGIDTRNNVYVWFMNKKMHAISFLVNTANAILRANLOK 1620
QY 1621 GENPSHYGITAENHPLNLTQOOLSEVALMTTSDVLSICVIFAMSFPAFVFELOER 1680
Db 1621 GENPSHYGITAENHPLNLTQOOLSEVALMTTSDVLSICVIFAMSFPAFVFELOER 1680
QY 1681 VSKAHLQFISGVKPYIWLNSFWDMCNVYPATLVIIIFICFOQKSYVSTNLPVAL 1740
Db 1681 VSKAHLQFISGVKPYIWLNSFWDMCNVYPATLVIIIFICFOQKSYVSTNLPVAL 1740
QY 1741 LLLLGWSITPLMPASVFKIPSTAYVLTISVNLFGINSVATFVLELFTDNKLNIN 1800
Db 1741 LLLLGWSITPLMPASVFKIPSTAYVLTISVNLFGINSVATFVLELFTDNKLNIN 1800
QY 1801 DILKSVFLIFPHECLGRGLIDMVKNQAMADALERFGENREVSPLSMDLVGNILPAMAVEG 1860
Db 1801 DILKSVFLIFPHECLGRGLIDMVKNQAMADALERFGENREVSPLSMDLVGNILPAMAVEG 1860
QY 1861 VVEFLITVLIQYRFFIRRPYNAKLSPLNDEDEVRRERORILIDGGQNDLLEIKELTKI 1920
Db 1861 VVEFLITVLIQYRFFIRRPYNAKLSPLNDEDEVRRERORILIDGGQNDLLEIKELTKI 1920
QY 1921 YRRKRKPAVDRLICVIGIPGECFGLGVNAGAKSSTFKMLTGDITVYTGDAFLNKNSTLSN 1980
Db 1921 YRRKRKPAVDRLICVIGIPGECFGLGVNAGAKSSTFKMLTGDITVYTGDAFLNKNSTLSN 1980
QY 1981 IHEVHONKGYCPOPDAITELLTGREHVEFFALLNGVBEKEVGKVAEIRKLGLVKYGEK 2040
Db 1981 IHEVHONKGYCPOPDAITELLTGREHVEFFALLNGVBEKEVGKVAEIRKLGLVKYGEK 2040
QY 2041 YAGNVSGNKKKLTSTAMALLGGPVVFLDEPTTGMDRKARFLMNCALSVKKEGRSVLT 2100
Db 2041 YAGNVSGNKKKLTSTAMALLGGPVVFLDEPTTGMDRKARFLMNCALSVKKEGRSVLT 2100
QY 2101 SHSMECEALCTRMAIWNNGFRCLIGSVOLHKNFGDGYITIVARIASNDLCPVODFFG 2160
Db 2101 SHSMECEALCTRMAIWNNGFRCLIGSVOLHKNFGDGYITIVARIASNDLCPVODFFG 2160

```

```

QY 2161 LAFPGSVLAKERRHNMLOYOLPSSLSLARIPLSOSKRRHLIEDYSVSQTTLDQVFN 2220
Db 2161 LAFPGSVLAKERRHNMLOYOLPSSLSLARIPLSOSKRRHLIEDYSVSQTTLDQVFN 2220
QY 2221 AKDQSDDDLKDLISLHKNQVAVVLTSLQDEKKESEV 2261
Db 2221 AKDQSDDDLKDLISLHKNQVAVVLTSLQDEKKESEV 2261

RESULT 13
AAB38115
ID AAB38115 standard; Protein; 2261 AA.
AC AAB38115;
XX
XX
DT 29-JAN-2001 (first entry)
XX
DE Human ABC1 cholesterol transporter mutant, R1587K.
XX
XX Human ABC1 cholesterol transporter; chromosome 9q31;
KW ATP-binding cassette; HDL deficiency disorder; high density lipoprotein;
KW Tangier disease; TD; familial HDL deficiency; FHA; polymorphism;
KW cardiovascular disease; coronary artery disease; coronary stenosis;
KW cerebrovascular disease; peripheral vascular disease;
KW Alzheimer's disease; Niemann-Pick disease; Huntington's disease;
KW X-linked adrenoleukodystrophy; cancer; gene therapy; genetic diagnosis;
KW prognosis; prophylaxis; drug screening; transgenic animal; mutant;
KW
XX
XX Homo sapiens.
XX
XX MO200055318-A2.
XX
PD 21-SEP-2000.
XX
PF 15-MAR-2000; 2000WO-1B00532.
XX
PR 15-MAR-1999; 99US-0124702.
PR 08-JUN-1999; 99US-0138048.
PR 17-JUN-1999; 99US-0139600.
PR 01-SEP-1999; 99US-0151977.
XX
XX (UYBR-) UNIV BRITISH COLUMBIA.
XX (XENO-) XENON BIOESPARCH INC.
XX
XX Hayden MR, Wilson AR, Pimstone SN;
XX
XX WPI; 2000-587528/55.
XX
XX New ABC1 polypeptide is useful for treating diseases associated with
XX ABC1 biological activity, e.g. Alzheimer's disease, Huntington's
XX disease and cancer.
XX
XX Examples; Page -: 229pp; English.
XX
XX The invention relates to the human ABC1 cholesterol transporter protein
XX (B38082) and to nucleic acid sequences (C69120) which encode it. ABC1 is
XX a member of the ATP-binding cassette (ABC transporter) superfamily of
XX proteins, and plays a crucial role in cholesterol transport, particularly
XX intracellular cholesterol trafficking in monocytes and fibroblasts, being
XX involved in cholesterol efflux from the cell. The gene encoding ABC1 is
XX located on chromosome 9q31, and mutations in this gene are associated
XX with two genetic HDL (high density lipoprotein) deficiency disorders,
XX Tangier disease (TD) and familial HDL deficiency (FHA). These diseases
XX are distinguishable in that TD is an autosomal recessive disorder, while
XX FHA is inherited as an autosomal dominant trait. Low levels of HDL ("good
XX cholesterol") in the blood correlate with a high risk of cardiovascular
XX disease, particularly coronary artery disease, but also cerebrovascular
XX disease, coronary stenosis, and peripheral vascular disease.
XX Conversely, a high level of HDL has protective effects against
XX cardiovascular disease. The invention provides genetic constructs and
XX transgenic cells and non-human animals comprising human ABC1 nucleic

```

CC acids, and methods of gene therapy for the treatment or prevention of
 CC cardiovascular disease comprising the administration of an expression
 CC vector encoding ABC1 or an active fragment thereof. The invention also
 CC encompasses compounds which mimic ABC1 activity, compounds which
 CC stimulate ABC1 expression and methods of screening for such compounds.
 CC It further relates to methods for determining whether a patient has an
 CC increased risk for cardiovascular disease due to polymorphisms in the
 CC ABC1 gene. Human ABC1 proteins and nucleotides can be used to treat
 CC or prevent cardiovascular disease, especially coronary artery disease,
 CC cerebrovascular disease, coronary restenosis or peripheral vascular
 CC disease. They may also be used in the treatment of diseases associated
 CC with ABC1 biological activity, such as Alzheimer's disease, Niemann-Pick
 CC disease, Huntington's disease, X-linked adrenoleukodystrophy and cancer.
 CC The invention specifically excludes proteins with the exact amino acid
 CC sequences of GenBank Accession No. CA10005.1 and X75926, and the nucleic
 CC acid with the exact sequence as GenBank Accession No. A012376.1. The
 CC present sequence represents a mutant human ABC1 cholesterol transporter
 CC associated with an altered cholesterol level and therefore an altered
 CC risk of cardiovascular disease.
 CC Note: The present sequence is not shown in the specification, but is
 CC derived from the native human ABC1 shown on pages 152-157.

XX Sequence 2261 AA:

Query Match 99.9%, Score 11786; DB 21; Length 2261;

Best Local Similarity 99.8%, Pred. No. 0;

Matches 2257; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 MACWPOLRLMKNTERRRQQLLEVAMPFLILISVRLSYPRYEDHECFPKA 60
 DB 1 MACWPOLRLMKNTERRRQQLLEVAMPFLILISVRLSYPRYEDHECFPKA 60
 QY 61 MSACATLPWVGIIICNANPCFRPTPGEPAGVGNFNKSVIARLFSAARLLYSQD 120
 DB 61 MSACATLPWVGIIICNANPCFRPTPGEPAGVGNFNKSVIARLFSAARLLYSQD 120
 QY 121 SKMDKRVKLTLOQIKSSSNLKLDFLVNDFSGFLYHNLSPKSTVDKMLRADVILH 180
 DB 121 SKMDKRVKLTLOQIKSSSNLKLDFLVNDFSGFLYHNLSPKSTVDKMLRADVILH 180
 QY 121 SKMDKRVKLTLOQIKSSSNLKLDFLVNDFSGFLYHNLSPKSTVDKMLRADVILH 180
 DB 121 SKMDKRVKLTLOQIKSSSNLKLDFLVNDFSGFLYHNLSPKSTVDKMLRADVILH 180
 QY 181 KVFLOGYQHLTSLCNGSKSEEMIOLOGDOVESLGLPKREKLAARLRSMIDLKFTL 240
 DB 181 KVFLOGYQHLTSLCNGSKSEEMIOLOGDOVESLGLPKREKLAARLRSMIDLKFTL 240
 QY 241 RLUNSTSPPSKELAEATKTLHSGLTLOELFSMRSDMDROEYFLTNVSSSSSTQI 300
 DB 241 RLUNSTSPPSKELAEATKTLHSGLTLOELFSMRSDMDROEYFLTNVSSSSSTQI 300
 QY 241 RLUNSTSPPSKELAEATKTLHSGLTLOELFSMRSDMDROEYFLTNVSSSSSTQI 300
 DB 241 RLUNSTSPPSKELAEATKTLHSGLTLOELFSMRSDMDROEYFLTNVSSSSSTQI 300
 QY 301 YQAVSRIVCGHEGGGLIKSLNMYEDNNYKALFGNGTEEDAEFTYDNSTPYCNDLMK 360
 DB 301 YQAVSRIVCGHEGGGLIKSLNMYEDNNYKALFGNGTEEDAEFTYDNSTPYCNDLMK 360
 QY 301 YQAVSRIVCGHEGGGLIKSLNMYEDNNYKALFGNGTEEDAEFTYDNSTPYCNDLMK 360
 DB 301 YQAVSRIVCGHEGGGLIKSLNMYEDNNYKALFGNGTEEDAEFTYDNSTPYCNDLMK 360
 QY 361 NLESSLRSITIKALKPLLVGKILYTPDTPATROYMAEYVNTFOGLAYFHOLEGMEELS 420
 DB 361 NLESSLRSITIKALKPLLVGKILYTPDTPATROYMAEYVNTFOGLAYFHOLEGMEELS 420
 QY 421 PLIWTFMENSQMDLYRMLDSRDHFMEQDLDGDMTAQDITVAFLLKHPEDVSSNGS 480
 DB 421 PLIWTFMENSQMDLYRMLDSRDHFMEQDLDGDMTAQDITVAFLLKHPEDVSSNGS 480
 QY 421 PLIWTFMENSQMDLYRMLDSRDHFMEQDLDGDMTAQDITVAFLLKHPEDVSSNGS 480
 DB 421 PLIWTFMENSQMDLYRMLDSRDHFMEQDLDGDMTAQDITVAFLLKHPEDVSSNGS 480
 QY 481 VYTWRAFNETQOAIKRTISRFMECVNLKLEPIATEVWLINSMELDERKFWAGIVETG 540
 DB 481 VYTWRAFNETQOAIKRTISRFMECVNLKLEPIATEVWLINSMELDERKFWAGIVETG 540
 QY 481 VYTWRAFNETQOAIKRTISRFMECVNLKLEPIATEVWLINSMELDERKFWAGIVETG 540
 DB 481 VYTWRAFNETQOAIKRTISRFMECVNLKLEPIATEVWLINSMELDERKFWAGIVETG 540
 QY 541 TTPGSTELPHNKKYKIRMDIDVEREKIKDGYWDPGPADEFEEMRYWGGFATLQOV 600
 DB 541 TTPGSTELPHNKKYKIRMDIDVEREKIKDGYWDPGPADEFEEMRYWGGFATLQOV 600
 QY 541 TTPGSTELPHNKKYKIRMDIDVEREKIKDGYWDPGPADEFEEMRYWGGFATLQOV 600
 DB 541 TTPGSTELPHNKKYKIRMDIDVEREKIKDGYWDPGPADEFEEMRYWGGFATLQOV 600
 QY 601 EDAIIRVLGTGKTKGVYQOAMPYCYDDIFLRVMSRSMPLFTLWYISAVVITIKGI 660
 DB 601 EDAIIRVLGTGKTKGVYQOAMPYCYDDIFLRVMSRSMPLFTLWYISAVVITIKGI 660
 QY 601 EDAIIRVLGTGKTKGVYQOAMPYCYDDIFLRVMSRSMPLFTLWYISAVVITIKGI 660
 DB 601 EDAIIRVLGTGKTKGVYQOAMPYCYDDIFLRVMSRSMPLFTLWYISAVVITIKGI 660
 QY 661 YEKEARKETMRIMGDINSILMFSWFISSLIEPLVASAGLLVLKLGILLPYSDRSVVFV 720

DB 661 YEKEARKETMRIMGDINSILMFSWFISSLIEPLVASAGLLVLKLGILLPYSDRSVVFV 720
 QY 721 FLVSFAVYVTLQCFLLISTLSFRANLAACGGITTFLLPYLVCAWQDYVGTLIKTPS 780
 DB 721 FLVSFAVYVTLQCFLLISTLSFRANLAACGGITTFLLPYLVCAWQDYVGTLIKTPS 780
 QY 781 LLSVAFGFCCEFALEEDGIGVQNDNLFESVFEEDGNLTTSISMLEDFDLGVMTM 840
 DB 781 LLSVAFGFCCEFALEEDGIGVQNDNLFESVFEEDGNLTTSISMLEDFDLGVMTM 840
 QY 841 YIAVFPQGYGIPRPWYFPCPKSYWGEESDEKSHGCSNOKRMSICEEPTHLKLGVS 900
 DB 841 YIAVFPQGYGIPRPWYFPCPKSYWGEESDEKSHGCSNOKRMSICEEPTHLKLGVS 900
 QY 901 IONLVKYYRQGMKVAVDGLANLFYEQGITSFLGNAGKTTMSILTGLFPPTSGATYL 960
 DB 901 IONLVKYYRQGMKVAVDGLANLFYEQGITSFLGNAGKTTMSILTGLFPPTSGATYL 960
 QY 961 GKDIRSEMTIRONLGYCPOHNVLPDMLTVEEHIMFYARLKGESEKHVAEMQMALDVG 1020
 DB 961 GKDIRSEMTIRONLGYCPOHNVLPDMLTVEEHIMFYARLKGESEKHVAEMQMALDVG 1020
 QY 1021 LPSSKLSKTSQLSGGMORRLSVALAFVGGSKVYLDEPTAGVDPYSRGIWELLKRYO 1080
 DB 1021 LPSSKLSKTSQLSGGMORRLSVALAFVGGSKVYLDEPTAGVDPYSRGIWELLKRYO 1080
 QY 1081 GRTTILSTHMDADVIGDRIAIISHGKLCQVSSSLFKNQLOGTGYLTLKKDVSSLS 1140
 DB 1081 GRTTILSTHMDADVIGDRIAIISHGKLCQVSSSLFKNQLOGTGYLTLKKDVSSLS 1140
 QY 1141 SCRRSSSTVSLKEDSVSSSDAGLSDESDTLIDVSAISNLRKHSARLVEDI 1200
 DB 1141 SCRRSSSTVSLKEDSVSSSDAGLSDESDTLIDVSAISNLRKHSARLVEDI 1200
 QY 1201 GHELTYVLPEAKEGAFVELFHEIDRLSDLGISYGISSETTLEETFLKVAEESVDAB 1260
 DB 1201 GHELTYVLPEAKEGAFVELFHEIDRLSDLGISYGISSETTLEETFLKVAEESVDAB 1260
 QY 1261 TSOGTLPARNRARAFGKQSCLRPFTEDDAADPNDSIDPESRETTDLISGMDKGSYQK 1320
 DB 1261 TSOGTLPARNRARAFGKQSCLRPFTEDDAADPNDSIDPESRETTDLISGMDKGSYQK 1320
 QY 1321 GWKLTQOQFVALLMKRLIIRSRKGFPAQIVLPAVEVCIALVFSLVPPGKYPSLEQ 1380
 DB 1321 GWKLTQOQFVALLMKRLIIRSRKGFPAQIVLPAVEVCIALVFSLVPPGKYPSLEQ 1380
 QY 1381 PWTNNEQYTVSNDAPEDTGTLELNLALYKDPGFGTRCMEGNPDPPTCOAGEEWTAP 1440
 DB 1381 PWTNNEQYTVSNDAPEDTGTLELNLALYKDPGFGTRCMEGNPDPPTCOAGEEWTAP 1440
 QY 1441 VPOTIMDLFQNGMTNMPNSPACCCSSDKIKKMLPYCPACGIPPRORONADILLQDL 1500
 DB 1441 VPOTIMDLFQNGMTNMPNSPACCCSSDKIKKMLPYCPACGIPPRORONADILLQDL 1500
 QY 1501 TGRNIDSYLVKTYVOIITAKSLKNTIWNFEFRYGGFSLGVSNTQALPPSOEVNDAIKOMK 1560
 DB 1501 TGRNIDSYLVKTYVOIITAKSLKNTIWNFEFRYGGFSLGVSNTQALPPSOEVNDAIKOMK 1560
 QY 1561 HUKLAKSSADRLNLSGREPTGLDTRNNYKWFENNGGMAISSPLAVINNAILRANLQK 1620
 DB 1561 HUKLAKSSADRLNLSGREPTGLDTRNNYKWFENNGGMAISSPLAVINNAILRANLQK 1620
 QY 1621 GENPSHYGITAFNHPNLTKOOLSEVALMTTSYDVLSICVIFRMSVVPASFVFLQER 1680
 DB 1621 GENPSHYGITAFNHPNLTKOOLSEVALMTTSYDVLSICVIFRMSVVPASFVFLQER 1680
 QY 1681 VSKAKHLOFISGVKPVYWLNSFPWDMCNVVPATLVITIFIGFOOKSYVSTNLPVAL 1740
 DB 1681 VSKAKHLOFISGVKPVYWLNSFPWDMCNVVPATLVITIFIGFOOKSYVSTNLPVAL 1740
 QY 1741 LLLYGSITPLAMPASFVKIPSTAVVLTSVULFINGSAVTPLYLEFTONKLANIN 1800

Db 1741 LLLXGMSITLMPASTVFKIPSTAYVLTSLVLTGINSVATFVLELFTDNKLNIN 1800

QY 1801 DILKSVFLIPHFCLGRGLDMVKNQAMADALERGENREVSPLSMDLVGRNLFAMA VEG 1860

Db 1801 DILKSVFLIPHFCLGRGLDMVKNQAMADALERGENREVSPLSMDLVGRNLFAMA VEG 1860

QY 1861 VVFLITVLTIOYRFFIRPRPVNAKLSPLNDEDEVREKORILDDGGQNDILEKELTKI 1920

Db 1861 VVFLITVLTIOYRFFIRPRPVNAKLSPLNDEDEVREKORILDDGGQNDILEKELTKI 1920

QY 1921 YRRKRKPAVDRIKVGIPGECFGLGVNAGAKSSTFKMLTGDITVTGDAFLKNKSLTN 1980

Db 1921 YRRKRKPAVDRIKVGIPGECFGLGVNAGAKSSTFKMLTGDITVTGDAFLKNKSLTN 1980

QY 1981 IHEVHONMGYCPQFDATITELLTGREHVEFALLRGVPEKEVKGWMAIRKGLVYKGEK 2040

Db 1981 IHEVHONMGYCPQFDATITELLTGREHVEFALLRGVPEKEVKGWMAIRKGLVYKGEK 2040

QY 2041 YAGNVSGNKKRKLSTAMALIGRPVVFLEPPTGMDPKARREFLMNCALSVYKEGRSVLT 2100

Db 2041 YAGNVSGNKKRKLSTAMALIGRPVVFLEPPTGMDPKARREFLMNCALSVYKEGRSVLT 2100

QY 2101 SHSMECEALCTRNAIMVNGRFCLGSVOHLKNRFGDGYTIIVARIAGSNPDLKPVDFFG 2160

Db 2101 SHSMECEALCTRNAIMVNGRFCLGSVOHLKNRFGDGYTIIVARIAGSNPDLKPVDFFG 2160

QY 2161 LAFPGSVLKEKHNMQLQYLPSSLSLARIFSLISQSKRHLIEDYVSQTTLDQVYVNF 2220

Db 2161 LAFPGSVLKEKHNMQLQYLPSSLSLARIFSLISQSKRHLIEDYVSQTTLDQVYVNF 2220

QY 2221 AKDOSDDHLKDLSLHKNOYVDVAVLTSFLQDEKVESYV 2261

Db 2221 AKDOSDDHLKDLSLHKNOYVDVAVLTSFLQDEKVESYV 2261

RESULT 14

AAB38105

ID AAB38105 standard; Protein: 2261 AA.

XX

AC AAB38105;

DT 29-JAN-2001 (first entry)

XX

DE Human ABC1 cholesterol transporter TD-2 mutant protein (Q597R).

XX

KW Human ABC1 cholesterol transporter; chromosome 9q31.

KW ATP-binding cassette, HDL deficiency disorder; high density lipoprotein;

KW Tangier disease; TD; familial HDL deficiency; FHD; polymorphism;

KW cardiovascular disease; coronary artery disease; coronary restenosis;

KW cerebrovascular disease; peripheral vascular disease;

KW Alzheimer's disease; Niemann-Pick disease; Huntington's disease;

KW X-linked adrenoleukodystrophy; cancer; gene therapy; genetic diagnosis;

KW prognosis; prophylaxis; drug screening; transgenic animal; mutant;

KW mutant.

OS Homo sapiens.

OS

XX

XX W020005318-A2.

XX

PD 21-SEP-2000.

XX

PF 15-MAR-2000; 2000WO-1B00532.

XX

PR 15-MAR-1999; 9905-0124702.

PR 08-JUN-1999; 9905-0138048.

PR 17-JUN-1999; 9905-0139600.

PR 01-SEP-1999; 9905-0151977.

XX

XX (UYBR-) UNIV BRITISH COLUMBIA.

PA (XENO-) XENON BIORESEARCH INC.

XX

PI Hayden MR, Wilson AR, Pimstone SN.

DR WPI: 2000-587528/55.

DR N-PSDB: AAC69386.

XX

PT New ABC1 polypeptide is useful for treating diseases associated with

PT ABC1 biological activity, e.g. Alzheimer's disease, Huntington's

PT disease and cancer -

XX

PS Examples: Page -: 229pp; English.

XX

CC The invention relates to the human ABC1 cholesterol transporter protein

CC (B38082) and to nucleic acid sequences (C69120) which encode it. ABC1 is

CC a member of the ATP-binding cassette (ABC transporter) superfamily of

CC proteins, and plays a crucial role in cholesterol transport, particularly

CC intracellular cholesterol trafficking in monocytes and fibroblasts, being

CC involved in cholesterol efflux from the cell. The gene encoding ABC1 is

CC located on chromosome 9q31, and mutations in this gene are associated

CC with two genetic HDL (high density lipoprotein) deficiency disorders,

CC Tangier disease (TD) and familial HDL deficiency (FHD). These diseases

CC are distinguishable in that TD is an autosomal recessive disorder, while

CC FHD is inherited as an autosomal dominant trait. Low levels of HDL ("good

CC cholesterol") in the blood correlate with a high risk of cardiovascular

CC disease, particularly coronary artery disease, but also cerebrovascular

CC disease, coronary restenosis, and peripheral vascular disease.

CC Conversely, a high level of HDL has protective effects against

CC cardiovascular disease. The invention provides genetic constructs and

CC transgenic cells and non-human animals comprising human ABC1 nucleic

CC acids, and methods of gene therapy for the treatment or prevention of

CC cardiovascular disease comprising the administration of an expression

CC vector encoding ABC1 or an active fragment thereof. The invention also

CC encompasses compounds which mimic ABC1 activity, compounds which

CC stimulate ABC1 expression and methods of screening for such compounds.

CC It further relates to methods for determining whether a patient has an

CC increased risk for cardiovascular disease due to polymorphisms in the

CC ABC1 gene. Human ABC1 proteins and nucleotides can be used to treat

CC or prevent cardiovascular disease, especially coronary artery disease,

CC cerebrovascular disease, coronary restenosis or peripheral vascular

CC disease. They may also be used in the treatment of diseases associated

CC with ABC1 biological activity, such as Alzheimer's disease, Niemann-Pick

CC disease, Huntington's disease, X-linked adrenoleukodystrophy and cancer.

CC The invention specifically excludes proteins with the exact amino acid

CC sequences of Genbank Accession No: CAA10005.1 and X75926, and the nucleic

CC acid with the exact sequence as Genbank Accession No: A012376.1. The

CC present sequence represents a mutant human ABC1 cholesterol transporter

CC associated with an altered cholesterol level and therefore an altered

CC risk of cardiovascular disease.

CC Note: The present sequence is not shown in the specification, but is

CC derived from the native human ABC1 shown on pages 152-157.

CC

XX

SQ Sequence 2261 AA;

Query Match 99.9%; Score 11785; DB 21; Length 2261;

Best Local Similarity 99.8%; Pred. No. 0;

Matches 2257; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 MACWPDLRLLLKMKNTLFRRRQTCQLLEAVMPLFILLISVRLSYPPYEDQHECFPKKA 60

Db 1 MACWPDLRLLLKMKNTLFRRRQTCQLLEAVMPLFILLISVRLSYPPYEDQHECFPKKA 60

QY 61 MPSAGTLPWVGITICNANNPCFRYPYTGEGARGVGNENFKSIVARLFSPARLLYSQKDT 120

Db 61 MPSAGTLPWVGITICNANNPCFRYPYTGEGARGVGNENFKSIVARLFSPARLLYSQKDT 120

QY 121 SMKDKRKVLRPTLQOIKKSSSNLKLDDPLVDNETFGSLTYHNSLPKSVVDMLRADYTLH 180

Db 121 SMKDKRKVLRPTLQOIKKSSSNLKLDDPLVDNETFGSLTYHNSLPKSVVDMLRADYTLH 180

QY 181 KVFLOGYQLHLTSLNGSKSEEMIQLDQEVSELGLPKREKLAARVLRNSNMDILKPIIL 240

Db 181 KVFLOGYQLHLTSLNGSKSEEMIQLDQEVSELGLPKREKLAARVLRNSNMDILKPIIL 240

QY 241 RFLNSTSPPSKEELAEARKITLHSLGTLAQELFSKRSMSDKRQVMTFTNNSSSSSTQI 300

Db 241 RFLNSTSPPSKEELAEARKITLHSLGTLAQELFSKRSMSDKRQVMTFTNNSSSSSTQI 300

QY 301 YQAVSRIVCGHEGGGLIKSLNMYEDNNYKALFGNGTEEDAETFYDNSTTPYCNIDLMK 360
 DB 301 YQAVSRIVCGHEGGGLIKSLNMYEDNNYKALFGNGTEEDAETFYDNSTTPYCNIDLMK 360
 QY 361 NLESSPLSRILWKALKPLLVGKILYPTDPATROYAVEVNTKFOELAVFHDLSEGNWEELS 420
 DB 361 NLESSPLSRILWKALKPLLVGKILYPTDPATROYAVEVNTKFOELAVFHDLSEGNWEELS 420
 QY 421 PKITFMENSOEMDLVRMLLDSRDNDHFWEQOGLDMDTADIDYAFIAKHEDEVOSSNGS 480
 DB 421 PKITFMENSOEMDLVRMLLDSRDNDHFWEQOGLDMDTADIDYAFIAKHEDEVOSSNGS 480
 QY 481 VYTWREAFENETQOALIRTSIRMECVNLKLEPIATEVNLINKSMELDERKFMAGIVTGT 540
 DB 481 VYTWREAFENETQOALIRTSIRMECVNLKLEPIATEVNLINKSMELDERKFMAGIVTGT 540
 QY 541 ITPGSIELPHHYKIKRMDIDNVERTNKIKDGYWDPGRADPFEDMRYVWGFAVLQDV 600
 DB 541 ITPGSIELPHHYKIKRMDIDNVERTNKIKDGYWDPGRADPFEDMRYVWGFAVLQDV 600
 QY 601 EGAIRVLGTGFKTGVMQMPYCYVDDIFLRVMSRSMPLMTLAMYAVAVIKIGIV 660
 DB 601 EGAIRVLGTGFKTGVMQMPYCYVDDIFLRVMSRSMPLMTLAMYAVAVIKIGIV 660
 QY 661 YEKEARLKETMRIMGLDNSILMFMSISLLPLVSAGLIVILKGLMLFYSDPSVYFV 720
 DB 661 YEKEARLKETMRIMGLDNSILMFMSISLLPLVSAGLIVILKGLMLFYSDPSVYFV 720
 QY 721 FLVSFAVYVILQCELLISTLFSRANLAACGGIYFTLYPYLVCAVADYVGYTLKIFAS 780
 DB 721 FLVSFAVYVILQCELLISTLFSRANLAACGGIYFTLYPYLVCAVADYVGYTLKIFAS 780
 QY 781 LLSPAFGGCGCYFALFEEOGIGVQMDLFBSPVEDEGNLTLSMMLPTOTFLYGVMTW 840
 DB 781 LLSPAFGGCGCYFALFEEOGIGVQMDLFBSPVEDEGNLTLSMMLPTOTFLYGVMTW 840
 QY 841 YLEAVFPQGYGIPRMWYFPCSTKSYWFGESDEKSHPGSNOKRMSICMEDEPTHLKGV 900
 DB 841 YLEAVFPQGYGIPRMWYFPCSTKSYWFGESDEKSHPGSNOKRMSICMEDEPTHLKGV 900
 QY 901 IONLVKVRDGGKAVAVDGLALNFYEGQITTSFLGHNGAGKTMTMSILTLGFPPTSGTAYIL 960
 DB 901 IONLVKVRDGGKAVAVDGLALNFYEGQITTSFLGHNGAGKTMTMSILTLGFPPTSGTAYIL 960
 QY 961 GKDISEMSTIKONLGVCPQHNVLFDMLTVEEHIWFAVRLGSLSHKAKEMQALDVG 1020
 DB 961 GKDISEMSTIKONLGVCPQHNVLFDMLTVEEHIWFAVRLGSLSHKAKEMQALDVG 1020
 QY 1021 LPSRSKLKSTQSLGSGMORKLSVALAFVGSKVVLIDEPAGVDPYSRRIWELLKRYQ 1080
 DB 1021 LPSRSKLKSTQSLGSGMORKLSVALAFVGSKVVLIDEPAGVDPYSRRIWELLKRYQ 1080
 QY 1081 GRTIILSTHMDADYLDGRIALISHKLCYGVSSLPFLKNOLGTGYVTLVKKDVESSL 1140
 DB 1081 GRTIILSTHMDADYLDGRIALISHKLCYGVSSLPFLKNOLGTGYVTLVKKDVESSL 1140
 QY 1141 SCRNSSSTVYLKEDSVSSQSSDAGLSDHSDTLTIDVSAISMLIKKHSEARLVEDI 1200
 DB 1141 SCRNSSSTVYLKEDSVSSQSSDAGLSDHSDTLTIDVSAISMLIKKHSEARLVEDI 1200
 QY 1201 GHELTYYLFEAAKGAFAVELFHEIDRLSDGISYGISETTLEIRFLKVAEESVDAE 1260
 DB 1201 GHELTYYLFEAAKGAFAVELFHEIDRLSDGISYGISETTLEIRFLKVAEESVDAE 1260
 QY 1261 TSDGTLPARNRARAFGKOSCLRPTEDDADPNDSIDDPESRETDLLSGMDGKSYOVK 1320
 DB 1261 TSDGTLPARNRARAFGKOSCLRPTEDDADPNDSIDDPESRETDLLSGMDGKSYOVK 1320
 QY 1321 GKKLQOQFVALLMKRLIARSRKGFPAQIVLPAPVCIALVSLVPPGKYPSLEQ 1380
 DB 1321 GKKLQOQFVALLMKRLIARSRKGFPAQIVLPAPVCIALVSLVPPGKYPSLEQ 1380

QY 1381 PMWNEQYFVNSDAPEDGTLELNLALTKDPGFGTCMEGNIPDPTQOAGEEEMWTAP 1440
 DB 1381 PMWNEQYFVNSDAPEDGTLELNLALTKDPGFGTCMEGNIPDPTQOAGEEEMWTAP 1440
 QY 1441 VPOTIMLPONGWNTMONBSPACOCSSDKTKMLPVCPPGAGGLPPPOKONTADILQDL 1500
 DB 1441 VPOTIMLPONGWNTMONBSPACOCSSDKTKMLPVCPPGAGGLPPPOKONTADILQDL 1500
 QY 1501 TGRNISDYLKTYVOILIAKSLKKNIWNEFRYSGFSLGVSNTALPPSOEVNDAIKOMK 1560
 DB 1501 TGRNISDYLKTYVOILIAKSLKKNIWNEFRYSGFSLGVSNTALPPSOEVNDAIKOMK 1560
 QY 1561 HLKAKDSSADRFNLNSLGRMTGIDTRNNYKVMFNKGMHAISPLNVLINNALIRANLQK 1620
 DB 1561 HLKAKDSSADRFNLNSLGRMTGIDTRNNYKVMFNKGMHAISPLNVLINNALIRANLQK 1620
 QY 1621 GENPSHGTAFNHPNLNLTQOQLEVALMTTSDVLVSLCVIFAMSFVPASFVFLQER 1680
 DB 1621 GENPSHGTAFNHPNLNLTQOQLEVALMTTSDVLVSLCVIFAMSFVPASFVFLQER 1680
 QY 1681 VSKAKHLOFISGVKPVLYMLSNFVMDKCNVVPATLVIIIFICEQOKSVYSSNTLPVL 1740
 DB 1681 VSKAKHLOFISGVKPVLYMLSNFVMDKCNVVPATLVIIIFICEQOKSVYSSNTLPVL 1740
 QY 1741 LLLLYGMSITPLMYPASFVEKIPSTAYVVLTVSNLFGINGSVATEVLELFTDNKLNIN 1800
 DB 1741 LLLLYGMSITPLMYPASFVEKIPSTAYVVLTVSNLFGINGSVATEVLELFTDNKLNIN 1800
 QY 1801 DILKSVFLIFPHCLGGLIDWYKNQAMADALEPFGENRFVSPSLSDLVGRNLFMAVAG 1860
 DB 1801 DILKSVFLIFPHCLGGLIDWYKNQAMADALEPFGENRFVSPSLSDLVGRNLFMAVAG 1860
 QY 1861 VVFLITVLIQYRFFIRPRVNAKLSPLNDEDEDVREERORILLDGGQNDILEIKELTKI 1920
 DB 1861 VVFLITVLIQYRFFIRPRVNAKLSPLNDEDEDVREERORILLDGGQNDILEIKELTKI 1920
 QY 1921 YRRKRKPAVDRIKICGIPGCEGFLGYNAGKSTFKMLGDTTVPDGAFLNKNLSILN 1980
 DB 1921 YRRKRKPAVDRIKICGIPGCEGFLGYNAGKSTFKMLGDTTVPDGAFLNKNLSILN 1980
 QY 1981 IHEVHQMKGCPQOPDATTELLTGREHVEFALLRGVBEKEVGVGEAATKGLVYGER 2040
 DB 1981 IHEVHQMKGCPQOPDATTELLTGREHVEFALLRGVBEKEVGVGEAATKGLVYGER 2040
 QY 2041 YACNYSGNKRKLIATAALLGGPPVFLDEPTTGMDPKARFPLMNCALSVYKGRSVLT 2100
 DB 2041 YACNYSGNKRKLIATAALLGGPPVFLDEPTTGMDPKARFPLMNCALSVYKGRSVLT 2100
 QY 2101 SHSMECEALCTRNAIMVNGRFRCIGSVQHLKKNFGDGYTIVRIAGSNPDLKPVODFFG 2160
 DB 2101 SHSMECEALCTRNAIMVNGRFRCIGSVQHLKKNFGDGYTIVRIAGSNPDLKPVODFFG 2160
 QY 2161 LAFPGSVLKEKHRMLOYOLPSSLSLARIFSLISQSKKRLHIEDVSVQTTLDQYFVNF 2220
 DB 2161 LAFPGSVLKEKHRMLOYOLPSSLSLARIFSLISQSKKRLHIEDVSVQTTLDQYFVNF 2220
 QY 2221 AKDOSDDHLKDLSLHKNOTYVDAVLTSLFLODEKKEVESYV 2261
 DB 2221 AKDOSDDHLKDLSLHKNOTYVDAVLTSLFLODEKKEVESYV 2261
 RESULT 15
 AAB38110
 ID AAB38110 standard; Protein; 2261 AA.
 XX
 AC AAB38110;
 XX
 DT 29-JAN-2001 (first entry)
 XX
 DE Human ABC1 cholesterol transporter mutant; V399A.
 XX
 KW Human ABC1 cholesterol transporter; chromosome 9q31;
 ATP-binding cassette; HDL deficiency disorder; high density lipoprotein;

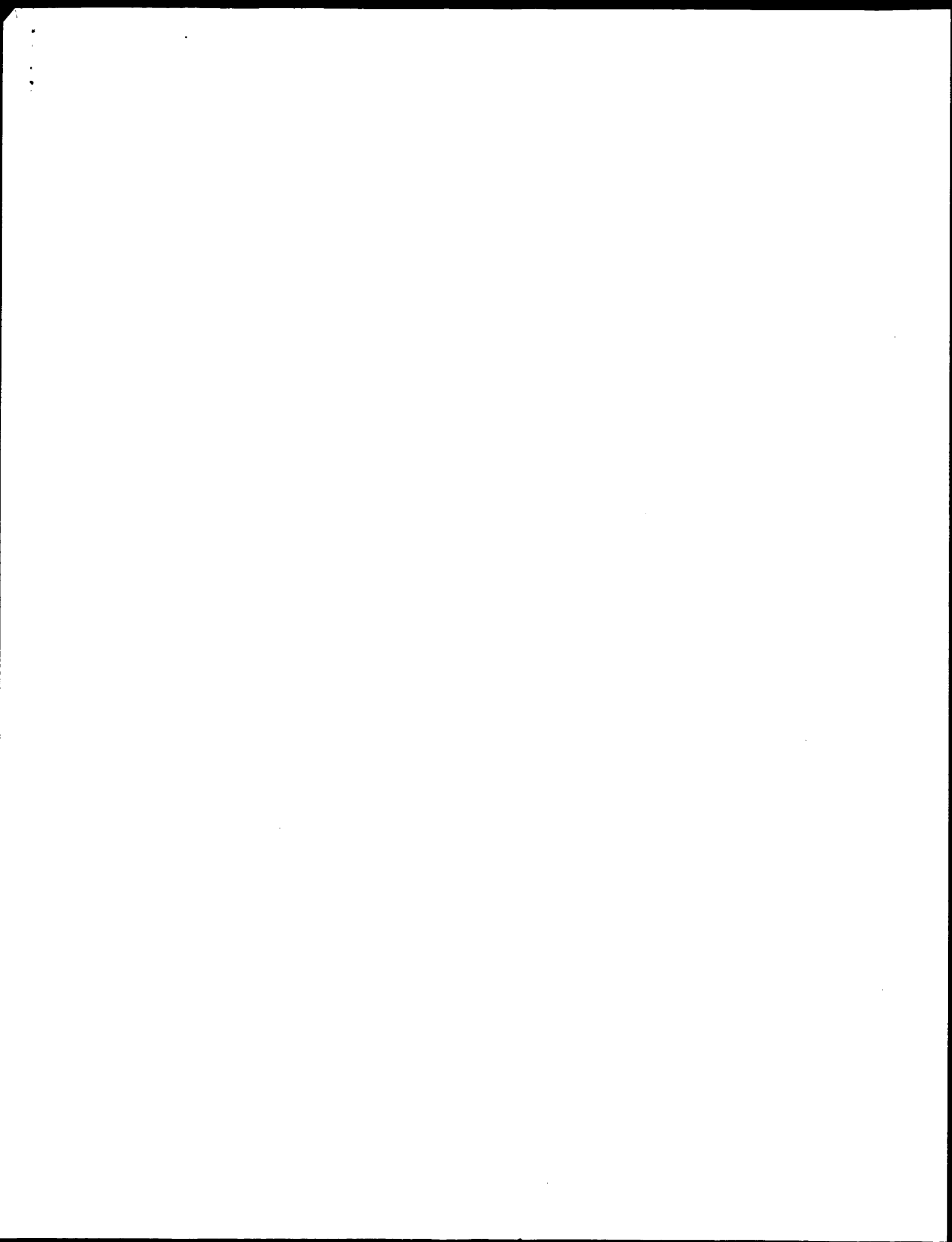
KW Tangier disease; TD; familial HDL deficiency; FHA; polymorphism;
 KW cardiovascular disease; coronary artery disease; coronary restenosis;
 KW cerebrovascular disease; peripheral vascular disease;
 KW Alzheimer's disease; Niemann-Pick disease; Huntington's disease;
 KW X-linked adrenoleukodystrophy; cancer; gene therapy; genetic diagnosis;
 KW prognosis; prophylaxis; drug screening; transgenic animal; mutant;
 KW mutant.
 XX
 OS Homo sapiens.
 PN WO20005318-A2.
 PD 21-SEP-2000.
 PF 15-MAR-2000; 2000WO-1B00532.
 PR 15-MAR-1999; 99US-0124702.
 PR 08-JUN-1999; 99US-0138048.
 PR 17-JUN-1999; 99US-0139600.
 PR 01-SEP-1999; 99US-0151977.
 XX
 PA (UYBR-) UNIV BRITISH COLUMBIA.
 PA (XENO-) XENON BIORESEARCH INC.
 PI Hayden MR, Wilson AR, Plimstone SN;
 XX WPI; 2000-587528/55.
 DR New ABC1 polypeptide is useful for treating diseases associated with
 PT ABC1 biological activity, e.g. Alzheimer's disease, Huntington's
 PT disease and cancer -
 XX
 PS Examples: Page -: 229p; English.
 XX
 CC The invention relates to the human ABC1 cholesterol transporter protein
 CC (B38082) and to nucleic acid sequences (C69120) which encode it. ABC1 is
 CC a member of the ATP-binding cassette (ABC transporter) superfamily of
 CC proteins, and plays a crucial role in cholesterol transport, particularly
 CC intracellular cholesterol trafficking in monocytes and fibroblasts, being
 CC involved in cholesterol efflux from the cell. The gene encoding ABC1 is
 CC located on chromosome 9q31, and mutations in this gene are associated
 CC with two genetic HDL (high density lipoprotein) deficiency disorders,
 CC Tangier disease (TD) and familial HDL deficiency (FHD). These diseases
 CC are distinguishable in that TD is an autosomal recessive disorder, while
 CC FHD is inherited as an autosomal dominant trait. Low levels of HDL ("good
 CC cholesterol") in the blood correlate with a high risk of cardiovascular
 CC disease, particularly coronary artery disease, but also cerebrovascular
 CC disease, coronary restenosis and peripheral vascular disease.
 CC Conversely, a high level of HDL has protective effects against
 CC cardiovascular disease. The invention provides genetic constructs and
 CC transgenic cells and non-human animals comprising human ABC1 nucleic
 CC acids, and methods of gene therapy for the treatment or prevention of
 CC cardiovascular disease comprising the administration of an expression
 CC vector encoding ABC1 or an active fragment thereof. The invention also
 CC encompasses compounds which mimic ABC1 activity, compounds which
 CC stimulate ABC1 expression and methods of screening for such compounds.
 CC It further relates to methods for determining whether a patient has an
 CC increased risk for cardiovascular disease due to polymorphisms in the
 CC ABC1 gene. Human ABC1 proteins and nucleotides can be used to treat
 CC or prevent cardiovascular disease, especially coronary artery disease,
 CC cerebrovascular disease, coronary restenosis or peripheral vascular
 CC disease. They may also be used in the treatment of diseases associated
 CC with ABC1 biological activity, such as Alzheimer's disease, Niemann-Pick
 CC disease, Huntington's disease, X-linked adrenoleukodystrophy and cancer.
 CC The invention specifically excludes proteins with the exact amino acid
 CC sequences of Genbank Accession No: CAA10005.1 and X75926, and the nucleic
 CC acid with the exact sequence as Genbank Accession No. A0102376.1. The
 CC present sequence represents a mutant human ABC1 cholesterol transporter
 CC associated with an altered cholesterol level and therefore an altered
 CC risk of cardiovascular disease.
 CC Note: The present sequence is not shown in the specification, but is
 CC derived from the native human ABC1 shown on pages 152-157.
 CC
 XX

Seq	Sequence	2261 AA:	99.98%	Score 11785;	DB 21;	Length 2261;
QY	1	MACPQRLILMKNTLFRROTCOLLLEVAMPFLFTLLISVSLSPVPEQHCPEPKNA	60			
DB	1	MACPQRLILMKNTLFRROTCOLLLEVAMPFLFTLLISVSLSPVPEQHCPEPKNA	60			
QY	61	MPASGTLPMVQGIICNANNPCFRYPGEGAPGVVGNENKSIYARLPSDARILLYSQKPT	120			
DB	61	MPASGTLPMVQGIICNANNPCFRYPGEGAPGVVGNENKSIYARLPSDARILLYSQKPT	120			
QY	121	SMKMRKVLRTLOQIKSSSNLKLQDFLVNDFPFGFLYHNLISLPKSTYDKMLRADYILH	180			
DB	121	SMKMRKVLRTLOQIKSSSNLKLQDFLVNDFPFGFLYHNLISLPKSTYDKMLRADYILH	180			
QY	181	KVFLQGYQLHTSLCNSKSKSEEMTOLGDOEVSLELGLPKREKLAARVYLRSMNDILKPTL	240			
DB	181	KVFLQGYQLHTSLCNSKSKSEEMTOLGDOEVSLELGLPKREKLAARVYLRSMNDILKPTL	240			
QY	241	RTLNSTSPFPSKELAEATKTLTSLGTLAQLFMSMSWSDMRQEVMTLVNVSSTSTQI	300			
DB	241	RTLNSTSPFPSKELAEATKTLTSLGTLAQLFMSMSWSDMRQEVMTLVNVSSTSTQI	300			
QY	301	YQAVSRIVCGHPGCGGLIKSLMWYEDNNYKALFGNGTGEDEAFETPDNSTTPYCDLKK	360			
DB	301	YQAVSRIVCGHPGCGGLIKSLMWYEDNNYKALFGNGTGEDEAFETPDNSTTPYCDLKK	360			
QY	361	NLSSPSLRITMKALKPLVGLKILYTPDPATRQVAEVAVKTPQELAVNHDLEGMEELS	420			
DB	361	NLSSPSLRITMKALKPLVGLKILYTPDPATRQVAEVAVKTPQELAVNHDLEGMEELS	420			
QY	421	PKIWTPEMSQENDVLMLLDSRDNDHFWEQQLDGLDWTADQIVAFIAHNPEDVSSNGS	480			
DB	421	PKIWTPEMSQENDVLMLLDSRDNDHFWEQQLDGLDWTADQIVAFIAHNPEDVSSNGS	480			
QY	481	VYTMREAFNETNOAIRISFMECVNLNKLDEPATVETWVILNKSMEILLDRKKFAGVIFPG	540			
DB	481	VYTMREAFNETNOAIRISFMECVNLNKLDEPATVETWVILNKSMEILLDRKKFAGVIFPG	540			
QY	541	ITPGSIELPHVKKYKIRMDIDNERTNKIKDGYWDPGRADPEDMRYVWGGFAYLDQVY	600			
DB	541	ITPGSIELPHVKKYKIRMDIDNERTNKIKDGYWDPGRADPEDMRYVWGGFAYLDQVY	600			
QY	601	EQATIRVLTGTEKKTGYVMOQMPYPCYVDIFLRVMSRSMPLFMTLAWITYSAVVIKGIY	660			
DB	601	EQATIRVLTGTEKKTGYVMOQMPYPCYVDIFLRVMSRSMPLFMTLAWITYSAVVIKGIY	660			
QY	661	YKPEARLKEFMRIAGDNLSTLMSWFTSSILPLVSAGLLVTLKGNLLPYSDPSVVFV	720			
DB	661	YKPEARLKEFMRIAGDNLSTLMSWFTSSILPLVSAGLLVTLKGNLLPYSDPSVVFV	720			
QY	721	FLSVFAVVTLLQCLFLISTLFSRANLAAACGGIITFTLLPYVLCAVADQYVGFYTLKIFAS	780			
DB	721	FLSVFAVVTLLQCLFLISTLFSRANLAAACGGIITFTLLPYVLCAVADQYVGFYTLKIFAS	780			
QY	781	LLSVNAGFCECEFAFLPEBGTGVOMDNLFESPVDEEDGNLTSTSMMLFDPPIVGVMTY	840			
DB	781	LLSVNAGFCECEFAFLPEBGTGVOMDNLFESPVDEEDGNLTSTSMMLFDPPIVGVMTY	840			
QY	841	YIYAVPFGQYGIIRPMYFPCTKSYWGESDESKSHSGSNOKRMSCEIMEEPTHLKLGVS	900			
DB	841	YIYAVPFGQYGIIRPMYFPCTKSYWGESDESKSHSGSNOKRMSCEIMEEPTHLKLGVS	900			
QY	901	IQMLVAVYRDGMKVAVDGALANFYEQQITSPFGHNGAGKTTMSITITGLPPTSGTAYTL	960			
DB	901	IQMLVAVYRDGMKVAVDGALANFYEQQITSPFGHNGAGKTTMSITITGLPPTSGTAYTL	960			
QY	961	GKDIRSEMSITRONLGVCPQHNVLFDMILVVEHIMVYARLKGISEKHVAEEMQALDVG	1020			
DB	961	GKDIRSEMSITRONLGVCPQHNVLFDMILVVEHIMVYARLKGISEKHVAEEMQALDVG	1020			

Qy	1021	LPSSKTKKTSQSLSGMORKLSVALFAPVGSKVILDEPTAAVDPSRGIWELLLEKXRO	1080
Db	1021	LPSSKTKKTSQSLSGMORKLSVALFAPVGSKVILDEPTAAVDPSRGIWELLLEKXRO	1080
Qy	1081	GRTTILSTHNHDEADVLDGRIATITSHGKLCQVSSLFKNQIGTGYTLTLKKKDVESLS	1140
* Db	1081	GRTTILSTHNHDEADVLDGRIATITSHGKLCQVSSLFKNQIGTGYTLTLKKKDVESLS	1140
Qy	1141	SRNNSSVSTVLKKEDSVSSSSSAGLGSFHEDTLITDVSAISNLKKHYSEARLYEDI	1200

[illegible]

```
Search completed: March  7, 2003, 08:47:08
Job time : 97 secs
```



Db 1 MAFWTOGLILMKNTYRRRQTFQLLIEVAMPLEFIILISVRLSYPRYEOHECFPNKA 60
QY 61 MSAGTLPWVOGICIANNPCTRYPPGAPAVGNFNKSIYARLEPSARILLYSOKDT 120
Db 61 MSAGTLPWVOGICIANNPCTRYPPGAPAVGNFNKSIYARLEPSARILLYSOKDT 120
QY 121 SMKDMKRYARTLOQIKSSSS-NIKLLODFVNDNETFSGLYHNLSPKSYVKMLRADYTL 179
Db 121 SMKDMKRYARTLOQIKSSSS-NIKLLODFVNDNETFSGLYHNLSPKSYVKMLRADYTL 179
QY 121 SIKDVOQVLAKIRKIGNSSGIDLKIRDLVNDNETSDPLRANVSNPSSAWEELDAEYVL 180
Db 121 SIKDVOQVLAKIRKIGNSSGIDLKIRDLVNDNETSDPLRANVSNPSSAWEELDAEYVL 180
QY 180 HXVFIQVLOLHLTSLCNSKSEEMTQGDQEV---SE---LQGLPEKELAAAEVRLSNMD 234
Db 181 QVYVSGRIQIRDLONCSNLSSEFLTIQNSVAMDSSEAFCTLPKETHLAELAFRANLN 240
QY 235 IKPLIRTNSTSPPEKSLAFATKTLHSLGTLAQLELFSNRSDMDQEVWFLTNVSS 294
Db 241 PLKPLQRELEFFNSL--KDLSTVVALDSDGLKLVKELLSKMSWSDMQEWFELTNVNS 298
QY 295 SSSQIYQAVSRITVCGHPEGGLKIKSLNWEEDNNYKALFGNGTEDEDAETFYDNSTPY 354
Db 299 NSSTQIYQAVSRITVCGHPEGGLKIKSLNWEEDNNYKALFGNGTEDEDAETFYDNSTPY 358
QY 335 CNDLKNLNESSPLSRIITKALKPLLVGLITLTPDTPATROYAEVNTFQELAVFHDEG 414
Db 339 CNELMKNLESSPLSRIIMKALKPLLVGLITLTPDTPATROYAEVNTFQELAVFHDEG 418
QY 415 MWEELSPKIMTFEMNSQMDLVRLMLDSRDNDFHEQOGLDGLMADODIVAFLEKHEBDV 474
Db 419 MWEELSPKIMTFEMNSQMDLVRLMLDSRDNDFHEQOGLDGLMADODIVAFLEKHEBDV 475
QY 475 OSSNSVYTWBENETNOAITRITISREMECVNLKLEPIATEVWLJNSKSMELDERKFWA 534
Db 476 EADNOMVYTWBENETNOAITRITISREMECVNLKLEPIATEVWLJNSKSMELDERKFWA 535
QY 535 GIVEFGITPGSTIEDPHHKKYKIRMDINVERTNKIKDGYMDPGRRAPDEPMKRYWGGFA 594
Db 536 GAVEFELIAPNSTELPQHKYKIRMDINVERTNKIKDGYMDPGRRAPDEPMKRYWGGFT 595
QY 595 YLQDVVEOAIIRVLGTEKTKGYVMQMPYPCYVDIIFLRYMSRSMPLFMTLAWIYSVAV 654
Db 596 YLQDVVEOAIIRVLGTEKTKGYVMQMPYPCYVDIIFLRYMSRSMPLFMTLAWIYSVAV 655
QY 655 IIKGIVYERKARKETRMKIMGLDLSIMPSFISLLIPLLVSAGLLVILKLGNLPSYD 714
Db 656 IIKGIVYERKARKETRMKIMGLDLSIMPSFISLLIPLLVSAGLLVILKLGNLPSYD 715
QY 715 PSVVEFLSVFAVVTIIOCFELISTFLSFRANLAAACGIIYFTLYLPVLCVAMODVYGF 774
Db 716 PSVVEFLSVFAVVTIIOCFELISTFLSFRANLAAACGIIYFTLYLPVLCVAMODVYGF 775
QY 775 LKIFASLLSPVAFGCEYFALPEEOGIGVQMDNLFESPVEEDGFNLJTSISMFLPDTL 834
Db 776 LKIFASLLSPVAFGCEYFALPEEOGIGVQMDNLFESPVEEDGFNLJTSISMFLPDTL 835
QY 835 YGVMTWITEVFPQOYGIIPRPWYFPCPKSYWGEESDEKSHPSNOKRMEICMEEPH 894
Db 836 YGVMTWITEVFPQOYGIIPRPWYFPCPKSYWGEESDEKSHPSNOKRMEICMEEPH 895
QY 895 LKIGVSIQNLVKKYRDMKVAVDGLALNFYEQOITSEFLGNAGAKTJYMSIILGLPPTS 954
Db 896 LKIGVSIQNLVKKYRDMKVAVDGLALNFYEQOITSEFLGNAGAKTJYMSIILGLPPTS 955
QY 955 GTAVIILKOLIRSEKSTIROMLGYCPOHNVLFDMJVEBHITWEYARLKGISEKHVAKEMO 1014
Db 956 GTAVIILKOLIRSEKSTIROMLGYCPOHNVLFDMJVEBHITWEYARLKGISEKHVAKEMO 1015
QY 1015 MALDVLGPPSSKLSKTSOLSGMQRKLSVALAFVSGSKVYIIDEPTAGVDPYSRRIWEL 1074
Db 1016 MALDVLGPPSSKLSKTSOLSGMQRKLSVALAFVSGSKVYIIDEPTAGVDPYSRRIWEL 1074
QY 1075 LKRYROGRTIILSTHMDADVLDRIAIISHGKLCQVSSFLKNQOLGTGYLTVKKD 1134
Db 1075 LKRYROGRTIILSTHMDADVLDRIAIISHGKLCQVSSFLKNQOLGTGYLTVKKD 1134

QY 1135 VESSLSSCRNSSSTVSYLKKEDSVSSSDAGLGSHPESDTLTDVYSATISNLRKHVSEA 1194
Db 1135 VESSLSSCRNSSSTVSYLKKEDSVSSSDAGLGSHPESDTLTDVYSATISNLRKHVSEA 1194
QY 1195 RLVYDICHIELTYVLPYEAKEGAFVELFHEIDRLSDLGISSYGISSETTLEEFILKABE 1254
Db 1195 RLVYDICHIELTYVLPYEAKEGAFVELFHEIDRLSDLGISSYGISSETTLEEFILKABE 1254
QY 1255 SGVDAETSDCTLPARRRRRRFGDQSCIRPEYDDADAPDSDIDPESRETDLLSGMDK 1314
Db 1255 SGVDAETSDCTLPARRRRRRFGDQSCIRPEYDDADAPDSDIDPESRETDLLSGMDK 1314
QY 1315 GSYOVKGMKLSQOQFALLMKRLLIARRSKGFFAOIVLPAVVCATALFSLIVPPFGY 1374
Db 1315 GSYOVKGMKLSQOQFALLMKRLLIARRSKGFFAOIVLPAVVCATALFSLIVPPFGY 1374
QY 1375 PSLELOPMMYNEQYTFVSNDAPEDTGLELNAJTKDQPGTRCMGPNIPPTCOAGEE 1434
Db 1375 PSLELOPMMYNEQYTFVSNDAPEDTGLELNAJTKDQPGTRCMGPNIPPTCOAGEE 1434
QY 1435 EMTTAPVPTIMDLFONGMNTMONPSPACOCSSDKIKMLPVCOPGAGGLPPOROKONTA 1494
Db 1435 EMTTAPVPTIMDLFONGMNTMONPSPACOCSSDKIKMLPVCOPGAGGLPPOROKONTA 1494
QY 1495 DIIQDLTGRNISDYLVKTYVOIIAKSLKNKIVNEFRYGGFSLGVSNTQALPPSOEVNDA 1554
Db 1494 DIIQDLTGRNISDYLVKTYVOIIAKSLKNKIVNEFRYGGFSLGVSNTQALPPSOEVNDA 1553
QY 1555 IKQMKHKLAKSDSADRLNLSGRPMGLDTRNNYKVPFNKKGHAIJSSPLNTVNNAIL 1614
Db 1554 IKQMKHKLAKSDSADRLNLSGRPMGLDTRNNYKVPFNKKGHAIJSSPLNTVNNAIL 1613
QY 1615 RANLQGENPSSHGIIATFNHPLNLTQOQSEVALMTTSVDVLSICVIFAMFVASFVY 1674
Db 1614 RANLQGENPSSHGIIATFNHPLNLTQOQSEVALMTTSVDVLSICVIFAMFVASFVY 1673
QY 1675 FLIOERYSKAKHLQFISGKYRYTYMLSNFWMDCMYVVPATLVIIITFCQOQSVSSTN 1734
Db 1674 FLIOERYSKAKHLQFISGKYRYTYMLSNFWMDCMYVVPATLVIIITFCQOQSVSSTN 1733
QY 1735 LPLVALLLLLGWSITPLMYPASFEFKIPSTAYVVLVTSNLEFIGINGSVAFVLELFTDN 1794
Db 1734 LPLVALLLLLGWSITPLMYPASFEFKIPSTAYVVLVTSNLEFIGINGSVAFVLELFTDN 1793
QY 1795 KLNINDILKSVFLIPHEFCGLIDVKNQOMADALFERGENRFPSPLSMDLVGRNLF 1854
Db 1794 KLNINDILKSVFLIPHEFCGLIDVKNQOMADALFERGENRFPSPLSMDLVGRNLF 1853
QY 1855 AAVAGVVEFLITVLIQYRFFIRPPRYNAKLSPLNDEDEEDVRRRQRLIDGSGONDLEI 1914
Db 1854 AAVAGVVEFLITVLIQYRFFIRPPRYNAKLSPLNDEDEEDVRRRQRLIDGSGONDLEI 1913
QY 1915 KELTKIYRRKRPAYDRICVIGIPGECFGLGVNAGAKSSTFKMLTGTDTYTRDAPLNF 1974
Db 1914 KELTKIYRRKRPAYDRICVIGIPGECFGLGVNAGAKSSTFKMLTGTDTYTRDAPLNF 1973
QY 1975 NSIISNIEHVQNMKYCQOFDAITELTGRHVEFBALLRGVPEKXGKYGEMAIRLGL 2034
Db 1974 NSIISNIEHVQNMKYCQOFDAITELTGRHVEFBALLRGVPEKXGKYGEMAIRLGL 2033
QY 2035 VYGEKRYAGNYSGGKRLKSTAMALIGPPVYVLDDEPTGMDPKARFPLMNCALSVYKEG 2094
Db 2034 VYGEKRYAGNYSGGKRLKSTAMALIGPPVYVLDDEPTGMDPKARFPLMNCALSVYKEG 2093
QY 2095 RSVVLTISHMECEALCTRMAIMVNGRFLGSVOHUKNRFQGDYTYIVHJAGSNPDLK 2154
Db 2094 RSVVLTISHMECEALCTRMAIMVNGRFLGSVOHUKNRFQGDYTYIVHJAGSNPDLK 2153
QY 2155 VODFFGLAFPGSVYLKRRNNLOVQOLPSSUSLARITSIISQSKRLHIEDYVSQOTLD 2214
Db 2154 VODFFGLAFPGSVYLKRRNNLOVQOLPSSUSLARITSIISQSKRLHIEDYVSQOTLD 2213


```

QY 1429 CQAGEEMTAPYPTIMDLFONGNMTNPSACCCSDKIKKMLPVCPPAGCLPPQ 1488
DB 1454 C-INATSMKTPSPSPNTHLPQOKWTAHPSPSCSKREKLTMLPECPBGAGLPPQ 1512
QY 1489 RKONTADIIODLTGRNIDLYLVKYVOITAKSLKNKIMVNERRGESFSGNTQALPSP 1548
DB 1513 RQORSEVIEODLTNRNIDSLYKTYTYPALLRSSLKSKFWNEORIGGISTG-GKLAIPIS 1571
QY 1549 QEVNDAIKQKKHKLAKDSSADRLNSLGR-----FMTGLDTRNN 1589
DB 1572 GE-----ALVGFSLGGMNNSGCVTRERASKEMLDLKLHETTDN 1613
QY 1590 VKWFMNKGMAHATSSFLNYINNAALLRANLOKGENSHYITAPNHLNLTQOOLEVALM 1649
DB 1614 IKWFMNKGMAHATSSFLNYINNAALLRANLPDRDPEEYITISQPLNLTKEOLSDITVL 1673
QY 1650 TTSVDVLSICVIFAMSEVPASFEVLIOERYSKAKHLOFISGVKPEVYIWLSNFWMDCN 1709
DB 1674 TTSVDVAVACVIFAMSEVPASFEVLIOERYTKAKHLOFISGVSPTYIWLNLMDIMN 1733
QY 1710 YVPATLVITTHCFOQKSYVSTNIPVLALLLHYGWSITPLMYPASFEVKIPSTAVV 1769
DB 1734 YVASAGLVYIGFQKRAYTSPDNLPALVSLMLYGWAVIIMYMPASFEVSTAYVA 1793
QY 1770 LTVSNLFIGINSVATFELEFTDNK-LNNIDILKSVLIEPHICIGLGLDMKNQAM 1828
DB 1794 LGSANFIEINSATFEVLEENRNTLIRENAMUKLIVPHICIGLGLDLSAIV 1853
QY 1829 ADALEREGNRFVSPLSMDLVGRNLPMAMAVEGVFELLITVLYOREFIRPREVNAKLSPL 1888
DB 1854 TTVYAOFGCEEYSANPEQWDLIGKNLVAMAEVGVYELLTLIQHHPFLTRWIAEPAREPV 1913
QY 1889 NDEDEVREORRILIDGSGNDILEIKELTKTYRKKRPVADRICVIGIPGCGELGVN 1948
DB 1914 PBDDBVAERORVMSGGKKTDLKLTLETKYSSSSPAVDRLCVGRPGCEFLGVN 1973
QY 1949 GAGKSTFEMLTGDTTVTRKGAFLNKNLSLNIHEVHONMGCPQFDAITELTGREHVE 2008
DB 1974 GAGKSTFEMLTGDTTVTRKGAFLNKNLSLNIHEVHONMGCPQFDAITELTGREHVE 2033
QY 2009 PFALLRGVEKEVGKVAIKRLGVKGEKAYAGNSGKNRKLSTAMALIGPPVFL 2068
DB 2034 LVARLRGVSKIEKRYANWNGIOSLISLADRLAGTYSOGNKRKLSTALIDGCPDLLL 2093
QY 2069 DEPTGMDEKARPLMNCALSVYKEGRSVYLTSHEMCECALCTPMALMVNGRPFCLGSV 2128
DB 2094 DEPTGMDEKARPLMNCALSVYKEGRSVYLTSHEMCECALCTPMALMVNGRPFCLGSV 2153
QY 2129 QHLKMPGDDGYTVIYVRIAGSN---PDLKPYODFPGLAIPGSVLKEKHRNMLQYOLPSSL 2184
DB 2154 QHLKMPGDDGYTVIYVRIAGSN---PDLKPYODFPGLAIPGSVLKEKHRNMLQYOLPSSL 2211
QY 2185 SSLATFETSLSSOKKRLHEDYSVOTLLDOYFVFAKADSGDDHKLKOLSLH 2236
DB 2212 SSLATFETSLSSOKKRLHEDYSVOTLLDOYFVFAKADSGDDHKLKOLSLH 2260

```

RESULT 3

002698

ID 002698 PRELIMINARY: PRT: 2281 AA.

AC 002698: 01-JUL-1997 (TREMblrel. 04, Created)

DT 01-JUL-1997 (TREMblrel. 04, Last sequence update)

DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)

DE ABC transporter.

OS Bos taurus (Bovine).

OC Eukaryota, Metazoa, Chordata, Craniata, Vertebrata, Euteleostomi;

OC Mammalia, Eutheria, Cetartiodactyla, Ruminantia, Pecora, Bovidae;

OC Bovidae, Bovinae, Bos.

OX NCBI_TaxID=9913;

RN [1]

RP SEQUENCE FROM N.A.

```

RC TISSUE=RETINAL ROD CELL;
RX MEDLINE=97248596; PubMed=9092582;
RA Iilling M., Molloy L.L., Molloy R.S.;
RT "The 220-kDa rim protein of retinal rod outer segments is a member of
the ABC transporter superfamily."
RL J. Biol. Chem. 272:10303-10310(1997).
DR EMBL: 090126; AAC48716.1; -.
DR InterPro: IPR003593; AAA_ATPase.
DR InterPro: IPR003439; ABC_transport.
DR Pfam: PF00005; ABC_tran; 2.
DR Prodom: PD000006; ABC_transport; 2.
DR SMART: SM00382; AAA; 1.
DR TIGR: TIGR01257; rim_protein; 1.
KW ATP-binding.
SQ
SEQUENCE 2281 AA: 257228 MW: 710D404C98F7A079 CRC64:

```

Query Match 49.7%; Score 5864.5; DB 6; Length 2281;

Best Local Similarity 50.1%; Pred. No. 0;

Matches 1158; Conservative 373; Mismatches 642; Indels 139; Gaps 26;

```

QY 6 QRLRLKMLTFRRRQNCOLLLEVAMPFLFLILISVRLSYPEEQHECHFPKAMPASG 65
DB 6 QRLRLKMLTFRRRQNCOLLLEVAMPFLFLILISVRLSYPEEQHECHFPKAMPASG 65
QY 66 TLPPVQGIICNANPCRRYPTTPGEARGVGNFKSIYARLFSDARKLLYSOKOTSMKDM 125
DB 66 TLPPVQGIICNANPCRRYPTTPGEARGVGNFKSIYARLFSDARKLLYSOKOTSMKDM 125
QY 66 MLPWLGIFCNVNNPCFOSPTAGEPGLVSNYNNSTIARVYRDFQELMDAPESQHLQGV 125
DB 66 MLPWLGIFCNVNNPCFOSPTAGEPGLVSNYNNSTIARVYRDFQELMDAPESQHLQGV 125
QY 126 RKVARTLQOI-----KSSSNLKLQDLVNDNFTSGFLYHNLSLPKSYDVKMLRAD 176
DB 126 RKVARTLQOI-----KSSSNLKLQDLVNDNFTSGFLYHNLSLPKSYDVKMLRAD 176
QY 126 WRRLRTISQMLNTRMRPERIAGRGIRIREVLADEDEMTLFTFLYKNIGLSVSYVYLLVNSQ 185
DB 126 WRRLRTISQMLNTRMRPERIAGRGIRIREVLADEDEMTLFTFLYKNIGLSVSYVYLLVNSQ 185
QY 177 VILKHYLVQY-QLHLSL-S-CKSGSKSEMTQLGDQEVSE-----LGLPEKRLAAERVL 229
DB 177 VILKHYLVQY-QLHLSL-S-CKSGSKSEMTQLGDQEVSE-----LGLPEKRLAAERVL 229
QY 186 V-REOPFARGVPDLMKDIACSEALIERFLIPQRAAQVRGSLCSLSQGLQWMDTL 244
DB 186 V-REOPFARGVPDLMKDIACSEALIERFLIPQRAAQVRGSLCSLSQGLQWMDTL 244
QY 230 RSNMDILKPLRTLNSPPEPSKEIAETKTLTSHSGLTLAGELFSRSMV-----SDMR--- 282
DB 230 RSNMDILKPLRTLNSPPEPSKEIAETKTLTSHSGLTLAGELFSRSMV-----SDMR--- 282
QY 245 YANVDK-----LHVPRRLDS-----RSQGMNLSRGRILSDMSPRI 284
DB 245 YANVDK-----LHVPRRLDS-----RSQGMNLSRGRILSDMSPRI 284
QY 283 -----QEVNLT---NVNSSSQIYOAVSRIVCGEGGGLIKSLNMYEDNN 329
DB 283 -----QEVNLT---NVNSSSQIYOAVSRIVCGEGGGLIKSLNMYEDNN 329
QY 285 QEFTHRPSVODLWVTRPLVOTGPEPFGQMLGLDLCGYEGGGSVFSFNMVEDNN 344
DB 285 QEFTHRPSVODLWVTRPLVOTGPEPFGQMLGLDLCGYEGGGSVFSFNMVEDNN 344
QY 330 YKALFGNGEEDAEFTFYDNSTTPYCNDLMLKNLSESPSLRIKALKPLLYGKIYTPPT 389
DB 330 YKALFGNGEEDAEFTFYDNSTTPYCNDLMLKNLSESPSLRIKALKPLLYGKIYTPPT 389
QY 345 YKAFGLGIDSTRKDPISYDERTTFCNALQSLSPNPLTKIARAKAPLIMGKILTFPDS 404
DB 345 YKAFGLGIDSTRKDPISYDERTTFCNALQSLSPNPLTKIARAKAPLIMGKILTFPDS 404
QY 390 PATROVAEYNAKTFQELAVFHDLEGMMEELSPKIMTFEMENSOEMDLVRMLDSRDNDFW 449
DB 390 PATROVAEYNAKTFQELAVFHDLEGMMEELSPKIMTFEMENSOEMDLVRMLDSRDNDFW 449
QY 405 PATRILKANKNSFEELEERAKLVKWEYEGPQIWFEPDKSTOMSMIRTLLENPTYKAFW 464
DB 405 PATRILKANKNSFEELEERAKLVKWEYEGPQIWFEPDKSTOMSMIRTLLENPTYKAFW 464
QY 450 EQQLDGLDMTADQIVAFIAKHPEDVOSSNGSVYTWBEAFNETNOAIRTISRMECVNLK 509
DB 450 EQQLDGLDMTADQIVAFIAKHPEDVOSSNGSVYTWBEAFNETNOAIRTISRMECVNLK 509
QY 465 NROGEGITAEAVLNFLYNGPREQADVDNENMDINFDIRALRLANOYLECILDK 524
DB 465 NROGEGITAEAVLNFLYNGPREQADVDNENMDINFDIRALRLANOYLECILDK 524
QY 510 LEPLATEVWLINKSMELDERKFMAGIVFTGTPGSLERPHYKVKIRAMDINNVERTNKI 569
DB 510 LEPLATEVWLINKSMELDERKFMAGIVFTGTPGSLERPHYKVKIRAMDINNVERTNKI 569
QY 525 FESYDEFOQLTORALSLLEENRFAVYPRDMIPWTSLEPHVKTIRKINDIDIVEETNKI 584
DB 525 FESYDEFOQLTORALSLLEENRFAVYPRDMIPWTSLEPHVKTIRKINDIDIVEETNKI 584
QY 570 KDGYWDPGRPADPEEDMRVYVWGFAVLQDVQQAIRVLGTGTEKKTYVMQMPPCYVD 629
DB 570 KDGYWDPGRPADPEEDMRVYVWGFAVLQDVQQAIRVLGTGTEKKTYVMQMPPCYVD 629
QY 585 KDRYMSGPRADVEDERFYIWGFAVLQDVQQAIRVRSQAQGEVPRGIVLQDMPPCPVD 644
DB 585 KDRYMSGPRADVEDERFYIWGFAVLQDVQQAIRVRSQAQGEVPRGIVLQDMPPCPVD 644
QY 630 DIFLRVMSRMPFLPLAMTISYAVIIRKIVYEKARKLETRKINGLDSILMEFSLSS 689
DB 630 DIFLRVMSRMPFLPLAMTISYAVIIRKIVYEKARKLETRKINGLDSILMEFSLSS 689
QY 645 DSEMITILNRCFPLFVNLAMTISYVSMVTKSVILEKELKETLKNQGVSNRYIWCWFLDS 704
DB 645 DSEMITILNRCFPLFVNLAMTISYVSMVTKSVILEKELKETLKNQGVSNRYIWCWFLDS 704
QY 690 LIPULVAGLIVILKLGNIPLPSDPSVVEFVLSYAVVTIIOGFLISLFSRAMLAAC 749
DB 690 LIPULVAGLIVILKLGNIPLPSDPSVVEFVLSYAVVTIIOGFLISLFSRAMLAAC 749
QY 705 FSLMSMSICLITLFIHNGRIHNSNFIIFLFLDLASTATIMQCFLLSFRASLAAC 764
DB 705 FSLMSMSICLITLFIHNGRIHNSNFIIFLFLDLASTATIMQCFLLSFRASLAAC 764
QY 750 GGIITFTLVLYLVLCVAMODYVGFLLKIFASLSPVAFGEGCEYFALPEEGIGVQWMDL 809

```

Db	765	SGVITFLYLPHILLCFAMQDRIITADMKNVAVLSLSPAFGCTETLAFEDQGVGLQMSN	824
Qy	810	FESPDEEGFNLUTISISMILFDFLYGVWVYIEAVEPGOYGPJRPWYFPCOTSWFG--	867
Db	825	CNSPREDDEEFSLMSKMKMLDLAYGLGLAMVYLDQVPRGOYGRPLWYFLIOESYWLGE	884
Qy	868	-----EESDEKSHPGSNQ-----KRMSEICMEEPETHKLGVSLQNLVKKYRDSMKV	914
Db	885	GCSTREERALEKTEPTEITEEMDEPREPEGINDCFERELPGJLVPGVCYKNVLIPEYGR	944
Qy	915	AYDGLALFEGQITSLFLGNHGACKTMTSLTGLPPTSGATYLKCOLKRESEMITRON	974
Db	945	AVDRNITFYESQITAFGLGHNGAGTTLTSLMTGLLPTSTGVLYGKDIETNLDAIROS	1004
Qy	975	LGCVPHNVHFLFMYLVEEHIFWFAALKGLSEKHVKAEMQMALDVGJLPSKSLKSTOLS	1034
Db	1005	LQMCQOHNLFLPHLTVAEHILFLYQNLKCRSMDEQALQLEAMLEDTGL--HKRNEARDLS	1063
Qy	1035	GMQOKLSVALAFVGSKVYLIDERTACVDYRSRGIMELLKTRQGTITLSTHNDQA	1094
Db	1064	GGVQRKLSVALAFVGDACKVYUDEPTSGDVPYSRRSIMDLLKTRSGRTITLSTHNDQA	1123
Qy	1095	DVIGDRALITSHGKLCQVSSSLFLKNOJGTGYLLVYKDVSSLSQCRNSSSTVSYLKK	1154
Db	1124	DLIGDRALITISQRIYCSSTPYFLKNCSTGYLLVYLR--KTIOSQGRNATCSQSK	1182
Qy	1155	EDVSQSSSDAGISGHESDUTLITVSAISNLIRKHVSFARLVEDIGHETVYLPYEAK	1214
Db	1183	GSEVRCRPAQAEATTPROYLDG---DYNELLDWVHHNVHPRAKLVEICIGRDLFLPKNFK	1239
Qy	1215	EGAFVELFHEIDRLSDIGISSTYGETTELEFLFKVAEESVDATDSGTLPARRNRA	1274
Db	1240	QRAYASIFRELEETLADLISFGISIDTFLPEELFLKVTEDLDSGHLFACGTQOKKEN---	1296
Qy	1275	FGDKQSLRP-----FTEDDADPNDSIDIDPRESTDLISGMDGKGYQVQGMKLTOQO	1328
Db	1297	INLHPGSGSEKAGOTPOQSSSHPEBPANHPRGQPPRRERHGRSLNS-----GARLTVQH	1352
Qy	1329	FVALLMKRLLIARSKRGFFAQVILPAVFCVIALVLSLIVPRGKYPSLELOPMYINQY	1388
Db	1353	VOALLVRFQHTIRSKHDFLAQIVLPAITFVFLALMSLITPRGEXPALTIHPRMAYGOOY	1412
Qy	1389	TFVSNDAPEDTGYLELINALTKDQFECTRCMEGNNPIPDTPCOQAGEEWTPAPQYIMDL	1448
Db	1413	TFFSMDQPDSEWLSALADVLVNRKFGFNRCILKEBMLPEPFC--GNSSPMKTPVSPDYTHL	1471
Qy	1449	FQNGNWTMONSPSPACOSSDOKIKKMLVPCRPVGGGLPPORPKONTABIIDDLGRLNSDY	1508
Db	1472	LQOQKWADQDPSPCRCSTREKLTMLDECEBGAGGLPPORRIORSTETIIDLDRWVSDP	1531
Qy	1509	LKTYVVOITIAKSLNKLIVNVEFRYGGFSLVSNQALPPOEYNDAIKQMKHLKLAKDS	1568
Db	1532	LKTYTPALIRSLSKFEVNNQRRGGISVG---GKLPAEPNGEALV-----	1575
Qy	1569	SADRFNLISGR-----FMIGLDIRNNVKKVFNKKGMHAIISFLNVI	1609
Db	1576	---GLTSDLOGLMNVSGGPMTRERAAKEMKPAFLKOLETEDIKNVFNKKGMHAIISFLNVA	1632
Qy	1610	NNAILIRAILKQGENSHVGTAFANHPNLITLQOOLSEVALMTTSVDVLYSICVIFAMSEVP	1669
Db	1633	NHAILIRASLHKRKNDEEGITVVISOPILNTEQOLSEITVLITSDAVAYALCIVIFAMSEVP	1692
Qy	1670	ASEVVELLQOERYSKAKHLOETSGVSPVLYUWLSNEFWMDCMNTVVEPATVYIILFICQOQSY	1729
Db	1693	ASFVYLTLQERKNKAKHLOPVSQVSPITTYMLNLFMLDMNTVSALVYVIFIGFOKKAY	1752
Qy	1730	VSTNMLPVALLLILLYGWSITPLMTPASVEKEIPSTAVVLTSVMLFLGINGSVATPYLE	1789
Db	1753	TSEENPALVALLMLYGMVAVIPMMYPAFLFDPITSTAVVALSCANLFGINGSVATPYLE	1812
Qy	1790	LFTDNK--LNNINDILKSVFLFPPHPCLCRGLIDMKXNQAMDALEPGENRFVSPISMDL	1848

Db	1813	LEFNNTLRLINAMLRKLLIIFPHCLGGLDIALSQAIVDVVAQCFEARSNSNPQWDL	1872
Oy	1849	VGRNLFMAAEVGVFLITVLIOYREFIIRPRVNAKLSPLANDEDEDVREKORILLGGQ	1908
Db	1873	IGKNLMAAEGVYFLITLLIOYOFFFSKWTTEPAKEPIITDEDDVAEEORIIISGNK	1932
Oy	1909	NDLIEIKELTKYIRKRRKRPVADIKCYGIPPGDEFGILGYNAGKSTFRKMLTGDTYTG	1968
Db	1933	TDILRLNELTKVYSGTSSPAVDRLCVGVRGECFCGLISYNGAGKTTTKRLMLTGDAVSG	1992
Oy	1969	DAFLINKSLTINSIEHVONNGYCPQFPAITELITGREHVEFFALLRGVEPEKVGWEMA	2028
Db	1993	DATVAGKSLINISIDVQSMGCPQFPAIDDLTGREHLIYVARLGVAEEIEBVTWMS	2052
Oy	2029	IRKLGLVYGEKVGANYSGCKNRKRLSTAMALICGPPVYFLDEPTTMDKARPLMNCAL	2088
Db	2053	IQSLGLSLYADRLAGTYSGCKNRKRLSTALAILGCPPLVLLDEPTTMDQARRMLNTIM	2112
Oy	2089	SVYKGRSVLTSISMEECEALCTRMALIMVNGRRCIGSVQHLKKNFGGTYIVVRIAGS	2146
Db	2113	GIIRBEERAVLTSISMEECEALCTRLIMVKGAFQCGITQHLKSKFGGCIYVTKIRSP	2177
Oy	2149	N---PDLKPVDFEGFLAPCSVLTKEKRRNMLQYDLPSSLSLARIFSIISOSKKRLHE	2200
Db	2173	KDDLPLDGLPNDQFPGNCPSCVQREIRRYNTLQGVSS--SLAIRIFRLVSHKDSLIE	2230
Oy	2205	DYSVSQTTLDQYFVNFPAKDKSDDDHLKDSLH	2236
Db	2231	EYSVITQTTLDQYFVNFPAKQONE--TYDPLPLH	2259
RESULT 4			
ID	Q9BZC4	PRELIMINARY:	PRT: 2146 AA.
AC	Q9BZC4:		
DT	01-JUN-2001 (TREMBLrel. 17, Created)		
DT	01-JUN-2001 (TREMBLrel. 17, Last sequence update)		
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)		
DE	ABC transporter member 7.		
GN	ABCA7.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	MEDLINE=21328888; PubMed=11435699;		
RA	Broccardo C., Osorio J., Luciani M.-F., Schirml L.M., Prades C.,		
RA	Shulein S., Anould I., Naudin L., Lafargue C., Rostler M., Jordan B.		
RA	Matei M.G., Dean M., Deneffe P., Chimin G.;		
RT	"Comparative analysis of the promoter structure and genomic		
RT	organization of the human and mouse ABCA7 gene encoding a novel ABCA		
RT	transporter.";		
RL	Cytogenet. Cell Genet. 92:264-270(2001).		
DR	EMBL: AF328787; AAK00959.1; "		
DR	InterPro: IPR003593; AAA_Atpase.		
DR	InterPro: IPR001899; Gram_pos_anchor.		
DR	Pfam: PF00005; ABC_tran; 2.		
DR	ProDom: PD000006; ABC_transportr; 2.		
DR	SMART: SM00382; AAA; 2.		
DR	PROSITE: PS00211; ABC_TRANSPORTER; UNKNOWN_1.		
DR	PROSITE: PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.		
DR	ATP-binding.		
SEQ	SEQUENCE 2146 AA: 234306 MW: 2391728055AD97E75 CRC64:		

Query Match	49.0%	Score 5775.5	DB 4	Length 2146
Best Local Similarity	50.3%	Pred. No. 0		
Matches 1148	Conservative 368	Mismatches 605	Indels 151	Gaps 26
Oy	1	MACPOLRLILKNLFRRCQCLILVAVMLFTLLISIVRSYSPYRQDCEHPFNKA	60	
	1			
	1			
	1			
	1			
Db	1	MAFQTLMLILKNMKNFRPQVQLLVETLWLFLEFLVLAVRSHPLLEHNECHFPNKR	60	

QY	61	MSASATLWVOCITICNANNPCFRYP7PGEAPGVGVGNFNKSIARVLFSPA RRLLYSOKDT	120
Db	61	LSASATVWMLQGLICNANNPCFOL7PGEPRGLSNFNDSLVSRLDARVYLGASHR	120
QY	121	SMKDRKVLFR7LQOIKRSSNKLQOEVLDNNEFSGFLYHNLISLPKSTVDKMLBAYTLH	160
Db	121	TLAGIGKILATLRARSTAO-----POPKOSPELPPML--	154
QY	181	KVFLQGIOLHLTSLCNGSKSEMIOLGDOEVSLEGLPEKRIAAEVRVLRNMDILKPTL	240
Db	155	-----DVAL-----LISL	164
QY	241	RLNSTSPFPEKLEAETKTLHLSIGTLAQELFSPRMSDMRQOEYMLTNNSSSTQI	300
Db	165	RT--EISGLALGQAOEPLSHLSAEADLAQELLARSLVEUR--ALLORRGSGLEL	219
QY	301	YQAVSRVYCGHPREGGGLIKLSLWYEDNNYKALFGCNGTEDEAFYDNSTPTYCNDLAK	360
Db	220	---LSEALCVRGSPSTVSGPSLWNYEASDMLVE---QPEPALPDSSISLSPCSLIG	272
QY	361	NLESPSLRIYMKALPBLVKGILITPTPATROYMAEVNTPQELAVFHDEBGMWELS	420
Db	273	ALDSHP7LSRLMRKLP7LIGKLEAPPTPKRLMAQVNTFEITLLRDVREVMENIG	332
QY	421	PK1TFEMENSQEMDVLVRMLDSRDNDFHEQOOLGDLPMTAQDIYAVFLAKHPEDVOS----	476
Db	333	PR1TFEMDNSSNVAMLQRLQMODBGR--RQRPGRGD-----HMEALRFLDP	379
QY	477	-SNGSYTMEAFNEJNOMAKIRISFMFCVNLKLEPIATEVW7INKSMELLDERKFVAG	535
Db	380	GSGG--YSMODAHADGHLVGLTVGVEECSLSDKLEAPSEALVSRALOLLAEHRFVAG	437
QY	536	IVTGTITPESIELPH-----HKYKIRMDIDNERTKIKDGVWDGDPADPEEDMR	587
Db	438	VFLPGEDESDPTHEPTDPLGPHVRIKIRMDIDVYTRTKIKDRWDGDPADP7LTDLR	497
QY	588	YVWGGFAY7ODVDBAII7VLTGTEKTKCYVMQOMP7PCYVDDITLRVNSRMP7LNTLA	647
Db	498	YVWGGFVY7ODV7EAA7VAV7SGANPRAGLY7LOOMP7PCYVDDVFLRVLSR7LPLFLTA	557
QY	648	W1YSVAV7IKGIVYKEKARLEK7MR7IGL7NLS7IMP7SS7IL7PL7VSA7GL7V7I7LK7G	707
Db	558	W1YSV7LV7K7AVR7EKET7RL7DM7VR7ANG7LS7RAV7ML7G7WEL7SIG7PL7LSA7L7V7I7K7G	617
QY	708	NLLP7SDS7VYV7PL7SV7AVV7I7IQC7L7S7LE7SRAN7LAAG7GL7IF7FL7Y7PY7LV7CAV	767
Db	618	DLLP7SHG7V7PL7FAA7VAV7YQ7SL7I7S7AF7FS7RAN7LAAG7GL7I7S7V7LY7LV7CAV	677
QY	768	QDYVGF7L7K7IAS7LSP7AF7EG7EY7AL7FEED7IG7VOM7DL7ESP7VED7G7NLT7S7IM	827
Db	678	RDR7L7PAG7R7VA7SL7LSP7AR7G7GES7L7AL7LEB7G7BDA7OH7N7GT7R7P7--ADVF7SLA7OY7S7L	736
QY	828	ML7DF7L7G7VMT7W7IE7AV7FQ7YQ7G7IP7RM7V7P7CT7S7YMF7GEB7ESDK7SH7PG7SNOK7RM7SE7IC	887
Db	737	L7L7D7AL7L7G7MT7W7IE7AV7G7OY7G7I7EP7NN7P7FR7RS7YMG7P7R7K7SAP7C7PT7LD--PKV7L	795
QY	888	MEED7P7HL7K7IAS7LON7LV7KY7Y7RDM7GV7VAD7GL7AL7N7FE7O7IT7S7FL7GH7G7AK7TT7MS7ILT	947
Db	796	VEEA7P7GL7SP7GV7SR7SL7EK7F7P7GS7P7O7AL7RG7SL7D7P7O7G7HT7AL7F7L7GH7G7AK7TT7S7IL7S	855
QY	948	GL7F7PT7SG7T7A7I7L7AK7D7R7SM7T7I7RON7IG7V7C7O7HN7V7L7DM7L7VE7EH7IM7F7AR7L7K7G7SE7KH	1007
Db	856	GL7F7P7SG7S7AF7I7L7GH7V7R7SS7MA7I7R7PL7H7C7P7OY7NV7L7DM7L7VE7H7W7F7G7R7L7K7G7LSA7V	915
QY	1008	VKA7ME7D7MAL7DV7G7P7SS7K7L7K7S7I7S7G7M7O7K7L7S7VAL7A7F7G7S7K7V7I7L7DE7P7AG7V7P7S	1067
Db	916	VGP7R7D7RL7L7OD7V7G7L--V7K7S7O7V7TR7H7LS7G7M7O7K7L7S7VAL7A7F7G7S7G7OY7V7I7L7DE7P7AG7V7P7S	974
QY	1068	R7G7I7W7ELL7L7K7Y7R7G7RT7I7L7S7H7MD7E7AD7V7L7G7R7I7AL7I7S7H7G7L7C7V7SS7L7K7LN7K7O7L7G7T7G7V	1127
Db	975	R7G7I7W7ELL7L7K7Y7R7G7RT7I7L7S7H7L7DE7AE7L7D7R7V7A7V7A7G7SL7CC7G7S7P7L7LR7H7L7D7S7G7V	1034

QY	1128	ITLVKDVESLSSCRNSSTSVYLKKEEDSVSSSSDGLGDHSEPTLIDVSAISNLI	1187
Db	1035	LTLLVAKRLPLTTN-----EKADTMEGSSVDRQCKKNGSGSVGTPPOLLAEBY	1082
QY	1188	RKHVEARLVJDICHELTYLPLPEAKBEGAYVELPHEIDORLSDGISSTLEI	1247
Db	1083	QHNVGATLVEELPHEILVLPRTGHDHSEFATLPEREDFELAEIRLTVGIGSDTSLEI	1142Z
QY	1248	FLKVAESGVDAETSDGTLPRARRNRARF-PKOSCLRPFEEDDAAPNOSDIDPRESREID	1306
Db	1143	FLKVVEECADTDMEDGSGOHLCTGAGLDVYLRILKMPQETALE--NGEPAGSAPETD	1200
QY	1307	LLSGDDKGSVOYKGMKLIQOOPVALIKMKRLIARRSRKGFFAOIVLPAVFCIALVESL	1366
Db	1201	OGSGPDVAG--RVQAGALLRQOLOALLKRLRLARRSRRLPFOIYLPALPFGVALVESL	1258
QY	1367	IVPFGKYPSLEILOPMYMEQYTFVSNDAEPDQTLLELMAALKDGFQTCRMEGNPLPD	1426
Db	1259	IVPFGHYPALRISPTMYCAQUSFFSEADAPDPRANLLPALLQOEFG-----LEE	1308
QY	1427	TPQOAGEEMTTAPVQUTIMDLFONGNNTWONPSAPACOSSDKIKAMIPVCPCGAGLPP	1486
Db	1309	PPVQSHSHRPSADEPAEYAKVLASGMTPEPSPAQCCSRQPARRLPDCPMAAGGPP	1368
QY	1487	PQRKONADILDLOGRNISDVLTYVOITAKSKKKIMVNEFRIGSGISVQNTQALP	1546
Db	1369	POAVTGSGEVYONLGRNLSDFLTKTYTRIVYKQGLTKKMYNEVRGEGSLG-GRDPGLP	1427
QY	1547	PSQEVDAIKOKKKHLKIAKOSSADRPFLNSIGRPMTGDLTNRNVVWPNNGGMAISSFL	1606
Db	1428	SGOELGRSVEEELMALSPDLPGALDRYLKMLTMAHISLDAQOSTKIWPNKKGWHSMAYFV	1487
QY	1607	NVIMNALILANLOKENGESHGIFAFPHPLNTLKOOLSPVALMTSVUVLSTVIFAMS	1666
Db	1488	NRASNALILAHLPBRPARHAISITTLNHPILNLTKEOULSBAALMASSVDVLSTCVFAMS	1547
QY	1667	FVPAASVVELIOERYSKAKHLQFTSGVKPYVYLNSFWMDMCMYVPAATVYIIFICFOO	1726
Db	1548	EVPASFTLVLEERYTRAKHLOLMGSLPTLYVLGNFLMDMONTLYPACIVLYLFLAQO	1607
QY	1727	KSYVSTNIPVYALLLILYGSTITPLMPASVYKIPSTAYVYLTSVNPLGINGSVATP	1786
Db	1608	RAYVAPANLPALLLLILYKGSTITPLMPASVYFSPSTAYVYLTCTNLETGINGSMAFP	1667
QY	1787	VELEPDKLNNINDIKSVLELFPFRCLGRGLIDMVANQOMADALRPFENRFPVSLM	1846
Db	1668	VELEFSDKLOEVSRIKQVFLFPFRCLGRGLIDMVANQOMADALRFLDRDROFQSLMW	1727
QY	1847	DLVGNLILPAMAVECVFELTYLQVRFIRRPVNAKLSPLNDEDVDVRRERORTLIDG	1906
Db	1728	EYVGKNLILAMIQOPLRLLTLLLOHRSOLLPOPVNRSPLLGEEDDAVAREBERVQGA	1787
QY	1907	GONDILELEKELTKYRRRRRPAVDRICVQIPRGEFCGLLGVNAGAKSTPFRWYTGDTLAS	1966
Db	1788	TQGDVLVLRNLTKYRQORMPAVDRICLGIPIRGEFCGLLGVNAGAKSTPFRWYTGDTLAS	1847
QY	1967	RODAVLKNSNLSINHEVYONHNGXOCPODATTELLTGREHVEFPALLRGVPEKEXGVE	2026
Db	1848	RGEAVLACHVADEPASAHLMSKGTCPQSDALFELLTGREHLEIARLYRPAEYAOVAG	1907
QY	2027	WAIKRLGIVKXERYAGNTSGGNKKRKLSTAMALLGCPVYVFLDEPTGMDPARREFLINC	2086
Db	1908	SGLARLGLISVYADRPAGTYSGGNKKRKLATLALYGDPAVYFLDEPTTGMDPARREFLINS	1967
QY	2087	ALSYVKGGRSVYLTISMSMECCALCTRPAIMVNGRFRCLGSVOHLKNRPDGYTVIVRA	2146
Db	1968	LLAYVKGGRSVYLTISMSMECCALCTRLAIWNGRFRCLGSPOHLKGRPAAGHTLTLRV	2027
QY	2147	GSNDLKPVOYDFGLGAPSVYLKEKHRIIMYOLO-PRSSLSLARFSTLSOSKRYLIED	2205
Db	2028	AARS--OPRAAFVAAPFESSSELREAHGGRRLRFLQRPQEGRCALARVFGELVANGAEHGED	2085
QY	2206	YXVSQTTIDQVYVNAKQDSDDHILKDLSTLKHQVY-VDVAV-----LTSFLADEKVK	2257


```

Db      2086  FSVSQTLLAEVFLFXSKDCKDE---DTEEQKEAGVGVDPRAGLQHPKRVSGFLDDPSTA 2142
QY      2258  ES 2259
Db      2143  ET 2144

```

RESULT 5	
Q9NR73	
ID Q9NR73	PRELIMINARY;
	PRT; 2146 AA

DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Macrophage ABC transporter.
GN ABCA7.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20334305; PubMed=10873640;
RA Kaminski W.E., Orso E., Diederich W., Klucken J., Drobnik W.,
RA Schmitz G.;
RT "Identification of a Novel Human Sterol-Sensitive ATP-Binding Cassette
RT Transporter (ABCA7)."
RL Biochem. Biophys. Res. Commun. 273:532-538(2000).
EMBL; AF250238; AAF85794.1; -;
DR InterPro: IPR003593; AAA_ATPase.
DR InterPro: IPR003439; ABC_transporter.
DR InterPro: IPR001899; Gram_pos_anchor.
DR Pfam: PF00005; ABC_tran; 2.
DR PRODom: PD000006; ABC_transporter; 2.
DR SMART: SM00382; AAA; 1.
DR PROSITE: PS00211; ABC_TRANSPORTER; UNKNOWN_1.
DR PROSITE: PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
DR ATP-binding. 2146 AA; 234469 MW; 799B16EB2D75FF0D CRC64;
SQ SEQUENCE

QY	421	PKIWTPEMNSOEJMDLRYMLDZSRONDHFMEOOLGDMTADDIYAFLAKPREVOVS----	476
Db	333	PRJFTEEMNDSNVAMLOJRLLOMDDGR--RQRPFGRD-----HMEALRSFLDP	379
QY	477	-SNGSVYTWREAFENENOAIKRTISRFMECVLNKLDEPIATEWVLIKSMELDERKFWAG	535
Db	380	GSOS--YSMODAIADVGHVLTGLRVTECLSDLKLEAPSEALVSRALQJLLEHRRPM	437
QY	536	IVFTGTPGSELPH-----HKYKIRMDIDNEVETNKIKGQYDPPRADPREDM	587
Db	438	VFLGPEDESSDPTHEPTPDLGPGHWRKIRMDIDVVTPTNKTIRQFMDPPRADPDLTDL	497
QY	588	YVMGFEYLIODVVEOALIRVLGTGTEKKGYVQOQWPCYCYDDJLEFRMYSRMLPEMTLA	647
Db	498	YVMGFEYLIODLVERAVALRVLSGAPNRACQJLQOQPYCYDDVDFEYRLXSRSLPFLITLA	557
QY	648	WISVAVILIGIYEEKEARLEKTRMTCIGDNSLWFSWFISSLPLVLVSAGLVLVLKG	707
Db	558	WISVVTLTVAAYVREKETRIJRDTRMRMGISRAVVLWGMFLSCLOGPFLLSALLVYLKLG	617
QY	708	NLEPYSDPSVVEVFLSVFAVVTILQCELI STLFSRANLAAACGCIYFTILYPLVLCVAM	767
Db	618	DILPSYSHQVVFLEFLAFAVAATVQSFLLSAFSSRANLAAACGIAVFSYLLPVLVCVAM	677
QY	768	ODVYGFLTKIFASLLSVNAGFCEXEFALFEEOGIGYOMDNLFESPDEDFNTTISIM	827
Db	678	RDRLPAGGRVAASLTSVNAFGFCESSLALLEBOGBOQMHNVGRPT--ADVFSLAOVSGL	736
QY	828	MLPFTFLYGWMTWLEAVFPFGQYGIJRPWYFPCQTSYVMEGSEDEKSHPGSNOKRMEIC	887
Db	737	LLDALYGLATWYLEAVCGQYGIJEPWNPFFRSRYMGCPRPSPAPCPTPLD--PKVL	795
QY	888	MEEPRTHLKLGVSIONLVKYYRDGKKVAVDGLALNFYEGQITSEFLGNHAGKTTTNSILT	947
Db	796	VEEAPRGISPGVSVRSLEKRFPSGPQALRGSLDIFYOCHITAFIGHAGAKTTTSLIS	855
QY	948	GLFPRTSGTATLIGKDIRSKMSTIRONLIGCOHNVLEPMLVEEHIMFYARLGLSEKH	1007
Db	856	GLFPSSGSAETLGHDRSSMAAIRPHLYGCOYVNLPEMLVVDHNVFVYRLGLSLAAV	915
QY	1008	VKAEMOQALDVLGPPSSKLSKTSOLSGMOKRLSVALAFVGSRYVLLDEPTAGVDPVS	1067
Db	916	VGPEDRDLQDVG--VSKOSVOJRHLSGGMOKRLSVALAFVGSQVYLDEPTAGVDPAS	974
QY	1068	RRGIWELLKRYROGRTIILSTHMDADVLGDRIALIISHGKLCYGVSSFLKNOLGTY	1127
Db	975	RRGIWELLKRYREGRITLITSTHMLDEAELLGDRVAAVAGRCLCCGSPFLRRHLSGTY	1034
QY	1128	LTVAKKOVESLSCRRSSSTSYLKKEDSVSSSSDADLGDHDSPTLTIIDVSAISLI	1187
Db	1035	LTVAKKARILPTTN-----EKADIDMGSVDTROEKKNSOGSRYGTPOLLALV	1082
QY	1188	RKHSEARLVEDIGHETLYVLYPEAAKEGAFVLEFIEDRLSDLGISYSGISETTEEI	1247
Db	1083	QHWVYGALVIEELHELVLVLYPTGAHDSFALLFREDLTRLAELRLTGYSDTSEI	1142
QY	1248	FLKVAEESGVDAEMSDGTLPARNRRAFG--DKOSCLRPTTEDDAADPNDSIDIPESRTD	1306
Db	1143	FLKVAEBCAADTDMEDGSCOHLCCTGJAGLDVTLRLKMPQOEFALE--NGEPAGSAPETD	1200
QY	1307	LLSGMDGGSYOVGKMLTQOOQVALLMKRLLLARSRKGFPAOVLVPAVFCIALVSL	1366
Db	1201	OGSGGDVAG--RVQGMALTRQOJLALLKREFLLARSRKGRLEQOVLVPAVFGIALVSL	1258
QY	1367	IYPPGKPYPLELQPMWINEQYTFVNSDADEDTGTELLNALTLPDGFEGTRCMEQNPID	1426
Db	1259	IYPPGCHPALRLSPTWYGAQVSPFSSDADGGRRLRLLEALLQENG-----LEE	1308
QY	1427	TPCQAGEEEMWTAPVOTIMDLFONGWMTWONSPACCCSSDKIKKMLVPCPGAGCJLP	1486
Db	1309	PEVQSSHSRFSAPVAPVAVALVSLGHWMTPESSPACCCSOPCARRLLPDCCPAAAGGPP	1368
QY	1487	PKRKONTADJLQDLTGRIISYLVKTYVQJIAASLKNKIMVNEFRGCGSLGVSMTQALP	1546

```

Db 1369 POAVTSGSEGVONTLGRNLSDFLVKTYPRVLVQGLTKKWNENYVGGFSLG-GMDPLP 1427
QY 1547 PQGEVNDATKQKKHKLAKDSSARFLNSIGREFTGIDTRNNKQVWPNNGMAISSFL 1606
Db 1428 SCQELGSRSEETLWALLSPGLGADRLVRLKNTLAWAHSLADDSIKIWNNGKSHMVAFV 1487
QY 1607 NVNNAILRANLQKGENPSHYGITAFAFNPLNLTQOOLSEVNLMTSVGLVSTCVIFAMS 1666
Db 1488 NKASAMAILRAHLPGRARAHNSITTLNPLNLTQOOLSEVNLMTSVGLVSTCVIFAMS 1547
QY 1667 FVPASVFPVFLIOERVSKAHLQFISGVKPVITWLSNFWDMCNVVPATVITIFIGEQ 1726
Db 1548 FVPASFTLVLEERVTAKHLQLMGSLPTLYLWGNFLMDCNVLPACIVVLEFLAQ 1607
QY 1727 KSYVSTNPLVATLLLLYGSSTIPMPAPASFEKIPSTAVVLTSVLFGINGSAVF 1786
Db 1608 RAYVAPALPALLLLLYGMSSTIPMPAPASFEKIPSTAVVLTSVLFGINGSMATF 1667
QY 1787 VLEFTDNKLNINDILKSVFLFPFHCLGRLIDMVNQMADALEFGEFNFPVPLSW 1846
Db 1668 VLEFSDQKLOEVSRLTKQVFLFPHCLGRLIDMVNQMADALEFGLDROFOSPLRW 1727
QY 1847 DLVGNLFMAVEGVFLLTVLIQYRFFTRPRVNAKLSPLNDEDEDVRRERQILDG 1906
Db 1728 EVVGKNTLAWYIQGPLFLFTLLQHSQILDPQPRVNSLPLGEBEDVAREFRVVOGA 1787
QY 1907 GONDILEIKELTKIYRRKRKPAVDRIQVGPGECPGLGNGAGKSTKMLTGDITV 1966
Db 1788 TQGDVLVRLNTLKYRGQRMPAVDRCLGIPGECFGLGNGAGKSTKRMVYGDITLAS 1847
QY 1967 RGDAFLNKSNTLSINHEVHQMGCYPOFDATITELLTGREHEFPALRGVPEKEVGKVE 2026
Db 1848 RGEAVLAGHSVAREPSAHLMSGYCPQSDALEFELLTGREHELLARLGRVPEAQVATAG 1907
QY 2027 WAIKRLGLVKGEYAGYSGNKRKLTSTAAALLGPRPVFLDEPTTGMPDKARFLMNC 2086
Db 1908 SGLARLGISWYADRPAGYSGNKRKLTATATLALVDPAVVFLEPTTGMPDSARFLMNS 1967
QY 2087 ALSVYKGRSVVLTSHMECEALCTRMATVNGRFRCLSVQHLKRRFEDGTIVYRIA 2146
Db 1968 LLAVVREGRSVMTLSHMECEALCSRLATVNGRFRCLSPQHLKRRFAGHTLIRVP 2027
QY 2147 GSNPDLKPVQDFGLAPPGSVLAKKHKNNMLQYOL-PSLSLARIETISLSQSKRLIHED 2205
Db 2028 AARS--OPAAAFVAAREPGELELREAHGRLRFQLPFGRCALARVGEFLAVHGAHEGVED 2085
QY 2206 YSVSQTLLDQVFNFAKQSDDDHLKDLSLHKQTY-VDAVAV-----LTSLQDEKVK 2257
Db 2086 FSVSQTMLEFVLFYSKQDQKDE---DTEQKQKAGVVDPAQLQHPKRVSOFLDPESTA 2142
QY 2258 ES 2259
Db 2143 ET 2144

```

```

RESULT 6
Q91V24 PRELIMINARY; PRT; 2159 AA.
AC Q91V24;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE ATP-binding cassette transporter sub-family A member 7.
GN ABCA7.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
NC NCBI_TaxID=10090;
RN [1]
RC SEQUENCE FROM N.A.
RP STRAIN=Ola129, and DBA/2;
RX MEDLINE=21328888; PubMed=11435699;

```

```

RA Broccardo C., Osorio J., Luciano M.-F., Schriml L.M., Prades C.,
RA Shulman S., Arnold I., Naudin L., Lafaque C., Rosier M., Jordan B.,
RA Mattei M.G., Dean M., Denelle P., Chini G.,
RT "Comparative analysis of the promoter structure and genomic
RT organization of the human and mouse ABCA7 gene encoding a novel ABCA
RT transporter."
RU Cytogenet. Cell Genet. 92:264-270(2001).
DR EMBL; AF287142; AAK56863.1; -.
DR EMBL; AF287141; AAK56862.1; -.
DR MGI; 1351646; Abca7.
DR InterPro; IPR003439; ABC_transporter.
DR InterPro; IPR002016; Peroxidase.
DR Pfam; PF00005; ABC_tran.2.
DR ProDom; PD000006; ABC_transporter; 2.
DR PROSITE; PS00211; ABC_TRANSPORTER; UNKNOWN_1.
DR PROSITE; PS00435; PEROXIDASE_1; UNKNOWN_1.
KW ATP-binding.
SQ SEQUENCE 2159 AA; 23682 MW; CD2BE3FE0DB8822B CRC64;

Query Match 48.6%; Score 5732; DB 11; Length 2159;
Best Local Similarity 50.6%; Pred. No. 0;
Matches 1144; Conservative 362; Mismatches 581; Indels 176; Gaps 31;

QY 1 MACPOLRLMLKNLFRRRQFCOLLLEVAMPFLFILLISVRLSYRPEDEHECPKA 60
Db 1 MALGTOLMLLKNNTYRRRQPIQLLVLLWFLFILLVAVRSHHPLEHHECHPKP 60
QY 61 MSAGTLPVVQGITCANNPCEFRYPPTGEPAGVGVNGFNKSIARLFSDARRLLYSQKDT 120
Db 61 LPSAGTVPWLOGVNCVNNSCFQHPPTGEPKPGVLSNFKQSLIRSLADPRTVL-----GGH 116
QY 121 SKMDMKVYRTQQLKSSSNLKIODELVNDNETFSGFLYHNLISLPKSYDKMLRADVILH 180
Db 117 STQMDLALGLKILPV----- 131
QY 181 KYFLQGYQLHLTSLNGSKSEEMIQLOEVSSELGLPREKLAARVL-RSNMDILKPI 239
Db 132 -----LNAVGGAGAPQE-----SDQPTSQ-GSVTKLL-EKLQKASLD---PV 169
QY 240 LRTLNSTSPFPEKLEALATKTLHSLGLTLOPLSMRSWSDMRQEFVFLTVNVSSTSQ 299
Db 170 L-----GQADSSMRKESDAIRDOELTLPLSIMELR--ALLRPPRGSGAGLE 215
QY 300 IYQAVSRIVC--GHPREGGLKTRSLNWEEDNNYKALFGNGTEDEATFTYDNTTPYCN 356
Db 216 L---VSEALCSTKQSSSGGL---SLNNYEANQLNEPFG---PEVALPDPNSLSPACS 265
QY 357 DLKMLESSPLSRIITWAKRLPLVGLKILYTPDPATQVMAEVNKTQDELAVFHDLEGMW 416
Db 266 EFVGTLDLDPVPSRLIMRRLKPLILGKIILFAPDTNPTKLMQVNOQTFELALLRDLHLM 325
QY 417 EELSPKWTFMENSOEMLVRLMLDSRNDHFWDQDLGDMTADQIYAFIAKRPEDVQS 476
Db 326 GVLGPIQIFENFMDSTNVMLDRLDLVGSTGQ--RQGPRAOKKLEAIKIDFL-----DP 376
QY 477 SNGSVYTWREAFENETNOAIRTISRFMECVNLKLEPIATEYWLINKSMELDERKEWAGI 536
Db 377 SRGG-YSMREAHADMGRLAGILGQWMECVSLDKLEAVSEBALVSRALELLEGRRLMAGI 435
QY 537 VFTG-----ITPGSTF---LPHVKKYKTRMDIDNVERTNKTIKDGYWDPGRADFEEDMY 588
Db 436 VFLSPHPLDSELSPPALSPGHILRFKIRMDIDVTYTRNKIRDFKMPGSPADPFMDLRY 495
QY 589 VMGGAAYLQDVAEQAIIHVLVTGTEKKGTVYVQOMPYCYVDITFLRWYSSMPLFMTLAW 648
Db 496 VMGGGVYIQLDLBOAANVAVLGGNSRGLYIQQPHHCYVDDVFLVRLSLSLPLETLAW 555
QY 649 IYSAVAILKGIYVEKAEKLKETMRIMGLDNLILWFSEISLPLPLVASAGLVVYIKGN 708
Db 556 IYSVALYVKAIVREKETRLRETMRAMGLSRAYVLMGLWFLSGLGFLVSAALVLYVILKGN 615
QY 709 LLPYSDBSVAVSVRAVVTITLQCFILSTLFSRANLAAACGGITTYTLYLVYLVCAWQ 768

```

```

Db 616 ILPYSHPVVLEFLAFAVAATVAQSFLLSAFSSRANLAAACGLAYFALYLPYLCVAMR 675
QY 769 DYVGFELKIFASLLSPVAFGCECFALFEEOGIGVOMDNLPESEVEDGFLNLTSSMM 828
Db 676 ERLHLGGLAASLLSPVAFGCECEALLLEEODGDAQOMHNLGTGA-EDVSLAOVSAFL 734
QY 829 LFDLFLYGVMTYEAVERPGYGIIPRMYPCTKSYWFGESDEKSHGSMOKRMSIEM 888
Db 735 LLDVAIVGLATWLEAVCGYGIPEPMNFPFRSYWCG-PGPRSSVLAAPADQPKVLY 793
QY 889 EEPFHKLKGVSIQNLVKKYRDMGVAVADGLALNFEQITSEFLCHNGAGKTTMSITG 948
Db 794 EEPFGLVGVASIRGLKHFRCPPALOGMLNDEEGHITAFIIOHNGAGKTTTSLISG 853
QY 949 LPPPTSGTAYILGKDIRSEMTITRONLGVCPQHNVLFDMLTVEEHIMWYARLKGSEKHV 1008
Db 854 LPPSSGASIIIGHVQTMALIRHLGICPOYNVFLMLYEEHVMYGRKLGSAAM 913
QY 1009 KAEOMALDVLPSKSKTSKTSOLSGQMOKLSVALAFVSGSKVYLIDEPTAGVPYSR 1068
Db 914 GERERLIDVGL-TLKRDTOTRHLSGMQRLSVALAFVSGSRVYIMDEPTAGVPSR 972
QY 1069 KGIWELLKYRGRTIITSTHMDADVLGRIALISHGKLCVSSSLFLKNOLGTGYL 1128
Db 973 KGIWELLKYRGRTIITSTHMDADVLGRIALISHGKLCVSSSLFLKNOLGTGYL 1032
QY 1129 TLVKKDVESLSSRNSSTVSYLKKEDSVSSQSSDAGLSDHESDT----- 1175
Db 1033 TLVK-----SSQSLYTHDAKGDESDPRREKSKSGMKTSDTAFTRGTSKSMQA 1081
QY 1176 -----LTIDVSAISNLRKHVSEARLVEDIGHLELYLVPEAKEGAEVLEFHEIDR 1228
Db 1082 PAPGAVPIPTSTARILELVQOHVPGAQLVEDLPHLELLVLPYAGALDSFAMVQELDQ 1141
QY 1229 LSDIGISSYGETLEEFELKVAEESGVADATSDGTPARANKRAFQDKOSCL-RPPT 1287
Db 1142 LELGLTGTGIDTLEELFELVDAHEGSDSRPOLHLR-----TCIPQPTG 1191
QY 1288 DDAADPNDSIDIPESRENDLISGMGKSGYQVKGMLTQOQFVALIMKRLILARRSGF 1347
Db 1192 PNASVLENGELAPQ-----GL-APNAQYQGWTLTCQOQALHLHRRFLARRSRGL 1242
QY 1348 PAQIYLPVAVFIALVFSILVPECKYSLELOPMWYNQYTFVSNDAPEDTGILELNA 1407
Db 1243 PAQVYLPVAVFIALVFSILVPECKYSLELOPMWYNQYTFVSNDAPEDTGILELNA 1302
QY 1408 LKRDGFGRCMEGNPIPTPCQAGEEEMTTPR-VPTIIMDFQNGNTMNPSPACCS 1466
Db 1303 LKRGGLQPSQMDKADARSECTHSLACFYVPEVPPVASTILASGNWTPESPACCS 1362
QY 1467 SDIKKMLPVCPPAGGLPPOPKONTADILDDTGRNISDVLVTKYQIILKSKNKTW 1526
Db 1363 QGARRLLPDCPAGAGPPRQAVAGLEGVONLGRNVSDPLVTKYISLVARGIKTKW 1422
QY 1527 VAEFRYGGSLGVSTQALPPEQEVNDATKOKKILKILAKODSADRTINSIGREMTGDT 1586
Db 1423 VAEFRYGGSLGVSTQALPPEQEVNDATKOKKILKILAKODSADRTINSIGREMTGDT 1481
QY 1587 RNNVYVFNKMGMAISSFLYINNAIIRANQKGENSHGITAIFNPIINTKQOLEV 1646
Db 1482 RNSLKIWNKMGMAISSFLYINNAIIRANQKGENSHGITAIFNPIINTKQOLEV 1541
QY 1647 ALMTTSVDLVISCIYFAMSPVAFVFLIOERYSKAKHLOFISGVPIVWLSNFWMD 1706
Db 1542 LTIASVDLVISCIYFAMSPVAFVFLIOERYSKAKHLOFISGVPIVWLSNFWMD 1601
QY 1707 MCNYYVPAIVLITFCQOKYSVSTMDPVALLLLYKMSITPPLMYPASFEKIPSTA 1766
Db 1603 MCNYYVPAIVLITFCQOKYSVSTMDPVALLLLYKMSITPPLMYPASFEKIPSTA 1661
QY 1767 YVVLISVNLFIGINSVATEVELFTDNKLNININDILKSVFLPHFPGTGLGIDMVNO 1826
Db 1662 YVVLISVNLFIGINSVATEVELFTDNKLNININDILKSVFLPHFPGTGLGIDMVNO 1721

```

```

QY 1827 AMADLERFCENREVPSPSLMDLVGRNLFAMAVEGVPELITVLIOYREFIRPPVNAKLS 1886
Db 1722 AMAAFERLFGKQFOSPLRMDIICKNLLAMAGCPFLITLLLOHNRNLLPOSKRRLP 1781
QY 1887 PLNDEDEDVRRERQRIIDGGQNDILEIKELTKYRRKRPAYDRICVGPPECEGLG 1946
Db 1782 PLGEDEDVAAQERBRVTKGATQGDVLRLDTKTVRCGRAPADRLCTGIPPECEGLG 1841
QY 1947 VNGAGKSSFTKMLTGDTTVTRGDAFLKNKNSILSIHIEVHONMGYCPQFDATITELTGREH 2006
Db 1842 VNGAGKSSFTKMLTGDTTVTRGDAFLKNKNSILSIHIEVHONMGYCPQFDATITELTGREH 1901
QY 2007 VEPFALLRGVEPEKVEGVGEWAIRKLGIVYGEKVEYACNYSKKRKLSTMALJGPPV 2066
Db 1902 LELFARLRGVEPEKVEGVGEWAIRKLGIVYGEKVEYACNYSKKRKLSTMALJGPPV 1961
QY 2067 FLDEPTTGMDPKARRPLMNCALSVKEGRSVLTSHSMECEALCTRMATVNGRFRCLG 2126
Db 1962 FLDEPTTGMDPKARRPLMNCALSVKEGRSVLTSHSMECEALCTRMATVNGRFRCLG 2021
QY 2127 SVQHLKRNFGDGTITVVRVLAGSNPDLKPVDFGLAFPGSVLKEKIRNMLQYOL-PSLS 2185
Db 2022 SSQHLKRGFGAGHRLTLRVDPDPE--PAIAFIRITPPGAELEHVSORLRLFOLPPGRC 2079
QY 2186 SLARFSLISQSKRKLHIEDSVSQTLDDQVFNFAKQDQDD 2228
Db 2080 TLTVFRELAAQGRNAHVEDFSQTLLEEVFLTFSKQGE 2122

RESULT 7
ID 096S58 PRELIMINARY; PRT; 2008 AA.
AC 096S58;
DT 01-DEC-2001 (TEMBLrel. 19, Created)
DT 01-DEC-2001 (TEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TEMBLrel. 21, Last annotation update)
DE ABCA-SSN.
GN ABCA7/ABCA-SSN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=2155283; PubMed=1135874;
RA Tanaka A., Ikeda Y., Abe-Dominie S., Atakawa R., Sadanami K.,
RA Kidera A., Nakagawa S., Nagase T., Aoki R., Kioka N., Amachi T.,
RA Yokoyama S., Ueda K.;
RT "Human ABCA1 Contains a Large Amino-Terminal Extracellular Domain
RT Homologous to an Epitope of Sjogren's Syndrome";
RL Biochem. Biophys. Res. Commun. 283:1019-1025(2001).
DR EMBL: AB055390; BAB62294.1; -.
DR InterPro: IPR003439; ABC_transprot.
DR InterPro: IPR001899; Gram_pos_anchor.
DR Pfam: PF00005; ABC_tran; 2.
DR Prodom: PD000006; ABC_transprot; 2.
DR PROSITE: PS00211; ABC_TRANSPORTER; UNKNOWN_1.
DR PROSITE: PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
SQ SEQUENCE 2008 AA: 218617 MW: 226F85C24230B90 CRC64;

Query Match 45.6%; Score 5376; DB 4; Length 2008;
Best Local Similarity 52.2%; Pred. No. 0;
Matches 1065; Conservative 334; Mismatches 554; Indels 88; Gaps 23;

QY 247 SPFSKELA-----EATKTLHSIGTLAQLFMSRSMKROEVMFLTVNNSSSSTQIY 301
Db 26 SPQSLGTLAQAOEPRLSHLSLEAEADLAQBLRLSLVELR---ALLQRPGRSGPLEL 81
QY 302 QAVSRIVCGHPEGGGLKIKSLNWEEDNNYKALFGNGNTEEDATETFDNSTTPPCNDLMKN 361
Db 82 --LSEALCSVRGSPSTVGPISLWMEASDLMELVG---QEPESALPDSSLSIPACSELIGA 135

```



```

QY 1500 LTRGRN1SDLVKTYOIIIAKSLKNKIWNEFRYGGFSGVSNTOALPPSQEVNDAIKOMK 1559
    |||
DB 1674 IIGHNVSXELTFSDF-----RLHRYGALTFG--NIQKSIPIAGICTRPLMVR 1720
    |||
QY 1560 KHLKLAKDSSADRLNSLGRFMTGIDTRNNKVMFNKGMHAISSFLAVINNAITRAKLO 1619
    |||
DB 1721 K-----IAVRVAQVLTNNKGHSMPTLYNSLNAITRAKLP 1757
    |||
QY 1620 KGE-NESHVGTAFNHPNLNLTQKOLSEVALMTTSVDVLSICIVFMSFVPSAFVFLIQ 1678
    |||
DB 1758 KSKGNPAAGVITVNNPMKTSASIS-LDYLLQGDVIAITFIYAMSFVPSAFVFLIVA 1816
    |||
QY 1679 EYVSAKHLQFISGVKPVYVLSNVMWMCNVVPAVLVIITFCFOQKSVYSTNLPLV 1738
    |||
DB 1817 EKSTPAKHLQFVGNPNVYVLMANVMDNLVLPATCCIIILFVFDLPATPTNPFAV 1876
    |||
QY 1739 ALLLLGKSTPLMYPASVFAFKIPSTAYVVLTSVNLFTGINSVATVLELFT--DNKLN 1797
    |||
DB 1877 LSLFLLIGKSTIPMYPASFWEVSSAVFLIVNLFTGITAIVATFLQLFEDKDK 1936
    |||
QY 1798 NINDIKSVELLFPHCLGRLIDMVKQAMADALEREGE--NRFVSPSLMDLVGNLFAM 1856
    |||
DB 1937 VVNSYLSKSCFLFPMYNGHGLMEIAYNEYAKIGQFDKMSPEWMDIVTRGIVAM 1996
    |||
QY 1857 AVEGVVFLITVLYRFRFIRPRVNAKLSPLNDEDEVRRERQILDDGGONDIETKE 1916
    |||
DB 1997 TVEGVGVFLITMCOYNEFROPRLPSTVKPEVD--DVVDASERQVLMGDADNDMVKLEN 2055
    |||
QY 1917 LTKIYRRK---RKPVNDRLCVGIPPEGCGFLGVAGKSSFTFKLTGDTVTGDAFLN 1973
    |||
DB 2056 LTKYKRSKRIIYAVDRICLGVBPGECEGLGVAGAKSTFTFKLTGDESBTGEAVN 2115
    |||
QY 1974 KSLISNHEVQNNQYCPORDATIELLTGHEHEFFALLGVPEKGVKGEVNAIRKLG 2033
    |||
DB 2116 GHISVTKDLQVQOSLGCPQDPAFLDELTAHEHQLDYLTRLGIPMKDQAVVMALEKLE 2175
    |||
QY 2034 LVKYEKAYAGNYSGNKPKLSTAMALIGPPVFLDEPTGMDPKARFELNCLASYKE 2093
    |||
DB 2176 LTKCADKPAAGSYSGNKKRSLTALALIGYPAFLDEPTGMDPKARFELNCLDIKT 2235
    |||
QY 2094 GRSVYLTSHSMEECALCTRAIYVNGFRCLGSYOHILKNRFGOGYITIVRIAGSNPDK 2153
    |||
DB 2236 GRSVYLTSHSMEECALCTRAIYVNGFRCLGSYOHILKNRFGOGYITIVRIAGSNPDK 2294
    |||
QY 2154 PYODEFFGLAFPGSVLEKEKRRMLQOLPSSLSLARIPSLISQSKRRHLIEDYSQTTL 2213
    |||
DB 2295 DTVREFNENFPEAMKEKHHKVOYLKSEHISLAQVFSKMEHVVGVLTEDYSVSOTTL 2354
    |||
QY 2214 DOYFVNFAKDOSDD 2227
    |||
DB 2355 DNVFVNFAKDOSDN 2368
    |||

```

```

RL Submitted (Aug-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX PubMed-11178988;
RA Kaminski W.E., Piehler A., Pullmann K., Porsch-Ozcurtun M., Duong C.,
RA Bared G.W., Buchler C., Schmitt G.;
RT "Complete Coding Sequence, Promoter Region, and Genomic Structure of
RT the Human ABCA2 Gene and Evidence for Sterol-Dependent Regulation in
RT Macrophages."
RL Biochem. Biophys. Res. Commun. 281:249-258(2001).
DR EMBL: AF178941; AAC09372.1; -.
DR EMBL: AF327657; AK1434.1; -.
DR InterPro: IPR003593; AAA_Alipase.
DR InterPro: IPR003439; ABC_transport.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR000572; Euk_Mb_oxred.
DR InterPro: IPR000566; Lipocin_cyFABP.
DR Pfam: PF00005; ABC_tran; 2.
DR Prodom: PD000006; ABC_transport; 2.
DR SMART: SM00382; AAA; 2.
DR PROSITE: PS00211; ABC_TRANSPORTER; UNKNOWN_1.
DR PROSITE: PS00022; EGF_1; UNKNOWN_1.
DR PROSITE: PS00213; LIPOCALIN; UNKNOWN_1.
DR PROSITE: PS00559; MOLYBDOPTERIN_EUK; UNKNOWN_1.
DR ATP-binding.
KW SQUENCE 2436 AA; 26995 MW; E044A3AF14EA25D1 CRC64;

Query Match 35.9%; Score 4240.5; DB 4; Length 2436;
Best local similarity 39.8%; Pred No. 26-269;
Matches 999; Conservative 347; Mismatches 729; Indels 435; Gaps 61;

QY 6 QLRLLKNTLFRROTQCLLEVAWPLFILLISVRLSYRPEDEQHECPNKA-MPSA 64
    |||
DB 6 QLRLLKNTLFRROTQCLLEVAWPLFILLISVRLSYRPEDEQHECPNKA-MPSA 65
    |||
QY 65 GLTPWVGIIICNANNCPFRPTGEAPGVGNPNKSIYARLSDSARLL---LYSQMDT 120
    |||
DB 66 GLTPWVGIIICNANNCPFRPTGEAPGVGNPNKSIYARLSDSARLL---LYSQMDT 115
    |||
QY 121 SMKDKRVLETLQOIKSSSNLKLQFLVDNETFSG-----FLYHNLSTP 165
    |||
DB 116 SLGSELEALR--OHLBALAGPQTSQSHDRSTVSFSLDSVARNPQELMRLTQNLSP 173
    |||
QY 166 KSTVDMKLADY---ILKRVFLQGLHTS--LNGS-----KSEEMI--- 204
    |||
DB 174 NSTAQLAARVDPPEVYHULFPGSSALDSQSLRKGQBPWSRLGNPLFRWEELLAPA 233
    |||
QY 205 ---QL-----GDQVSELCGLPKKELAAE-----RVLSNMDI 235
    |||
DB 234 LLEQLCTPGSGELGRLILVPEQKALQGYRDVAVCSGAARARFRSGLSALRNOLDV 293
    |||
QY 236 LKPIELTINSTPFSKELAE--TKTLHSLGTLAQLEFSKRSWDMKQEVMTLVNVS 293
    |||
DB 294 AK-VSQQLDLPAPNGSDSSPOAPPPRRLQALLGDLLD-----AKVIGLDVYLS 341
    |||
QY 294 SSSSTQIYQAVSRIVCGHPEGGLKIKSLMWYEDNNYKALRGNGTEEDAEFTYNSTTP 353
    |||
DB 342 ALALLLPQACGRTGPPASAGGAAN-----GTGAGVAMPNATFEGASAAATAP 396
    |||
QY 354 YCNLMKLNLESSLRIITKALKPLVG----- 381
    |||
DB 397 ---DPLQGGCSAFVQ--LWAGLQPIICGNRPITPEALRGMSSLGFTSKQORNLGLLY 451
    |||
QY 382 -----KTLVTPDTPATQVAAEVNKTFOELAVYHDLGEMIEELSPKIMFMENSQEMD 434
    |||
DB 452 HMTSNPKLITVAPAGEVDVRLKANETFAFGVNTHYAOWLNTSABIRSFLEQGRLOQ 511
    |||
QY 435 LVKMLDSDHNDHFWEQQLDGLDWTADQIVAFIAKHPEDVOSNOSVYTWTFEAFNETN-- 492
    |||
DB 512 HLRWL-----QOYVAFELRIHPF---ALNLSDELPPALRQDNFS 547
    |||
QY 493 -----QAITFTIS-----RFMECVNLKLPEDIAEVLINKSMF--LLDERKFWAG 535
    |||

```

RESULT 9
 Q9HC28 PRELIMINARY; PRT; 2436 AA.
 AC Q9HC28;
 DT 01-MAR-2001 (Tremblrel. 16, Created)
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE ATP-binding cassette sub-family A member 2 (ABC transporter
 DE ABCA2).
 GN ABCA2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=BRAIN;
 RA Vilevsk B., Chen Z., Davis W. Jr., Walsh E.S., Tew K.D.;
 "Cloning and characterization of human ABCA2.";
 RT

```

Db 548 LPSGALLQDITDIDNACGMIQFMKSVSDIPKGFPEDESVIVNTLNOAVQDNVFEAS 607
QY 536 IYFTITGSIELPHVYKTKRMIDINVERTNKIKDQWIDGCPADPREDMRYWGGFAY 595
Db 608 VIFQTRKSGS--LPEHVAHYKIRONSFFTEKTEINIRRAYWRGPRNTG--GRFYLYCGVW 662
QY 596 LODVVEQAIIRVLGTE--KKTGVYQMOQPCYVDIDFLRWSSRMPPLFMITAMITYSAV 654
Db 663 IODMMERAIIDTFVGHADVPEGSYQMPFPYCYTRDDFLFVIEHMPICAMISWYISAM 722
QY 655 IIKGIVYEKARKETMBRIMGLDINSIMFSWISSLLPLVLSAGILVYLKGLNLPYSD 714
Db 723 TIQHIAEKEHKLKEMTKMGLNNVHWAMFETGVLSTISVTLATLKYCGVLMSSH 782
QY 715 PSVIVVFLSVFVAVITLLOCFILSTLFSRANLAACGGIITVLYPYLC-----VAQD 769
Db 783 VVIIMFLAVAVATIMCFVLVSVLSAKKLASACGGIITVLYPYVVAIRREBVDK 842
QY 770 YVFLTKIFASILSPVAFGCEYFALFEEDGIGVQMDNLESPEEDGFLNLTISML 829
Db 843 ITAFI-KCIASIMSTTATGCGSKYFALYEVAGVGIQHTFSGSPYEGDPLNLLAVTLM 901
QY 830 FDFLYGVNTWYIEAVFPGQYIPRPWYFPCTKSYWGE--ESDEKSHSGSNOKRMS-- 884
Db 902 VDAVYVGLTWYIEAVHGMVGLPRPWYFPLQKSYWLGSGRTEAMEMSWPARFPLSV 961
QY 885 ---ETC-----MEEPHKLKGVSTONLVKYRQOMKAVAGCLANFVEGQIT 929
Db 962 EEDQACAMBSRFEETRGMEEPHPLPLVVCYDKLTLYKDKDKLTKLKLSTLNTENOV 1021
QY 930 SFLGHGAGKTTMSILGLFPPTSGTAYILGKDIRSESTROMIGVCPQHNVLFDMLT 989
Db 1022 SFLGHGAGKTTMSILGLFPPTSGTAYILGKDIRSESTROMIGVCPQHNVLFDMLT 1081
QY 990 VEBHIMFARLKLSEKHYKAEEMOALDVLPSKSLKSKTQSLSGQMRKLSVALAVG 1049
Db 1082 VEBHIMFARLKLSEKHYKAEEMOALDVLPSKSLKSKTQSLSGQMRKLSVALAVG 1140
QY 1050 GSKVYVLDPTAGVDPYSRGIMELLKRYROGTITLSTHMHDEADVIGDRATISHGKL 1109
Db 1141 GSRATILDEPTAGVDPYARRAATMDLTKKPGRTITLSTHMHDEADVIGDRATISHGKL 1200
QY 1110 CCVGSLLFKNOLGTYLTLVKKWDESSLSSCRNSSSTVSYLKKEDSVSSQSDAGLS 1169
Db 1201 KCCGSPFLKGTGYDGYRLTLVYKRAEPG-----GPGPGLAS 1238
QY 1170 DHESTLITDVS--AISNLIRKHAVFARVEDIGHELTVLVLEAKKBCAFELTHEIDD 1227
Db 1239 SPGRAPLSSCSSELOYSOFIRKHVASCLDSTSTELSYLSEAAKGAERLFOHLER 1298
QY 1228 RLSDLGISYVGISETTLEEFILKVAEE-----SGVDA-ETSDGTLR----- 1267
Db 1299 SLDALHLSFGMLDITLLEEVFLKVSFEEDOSLENSADVYESKDVLPAGAGSGBGHAG 1358
QY 1268 -----ARRNRA-FGDKOSCLRPTEDDAADPNDSDIDPESR 1303
Db 1359 NLARCELTQSOASLQSSASVSARGDEAGYTDVYGDYRPLF--DMPQCPD--NWSLQGV 1415
QY 1304 EYDLSGMDGKGYQYKMKLQOQFVALLMKRLIARSKRGFAQIYLVPAVPCIALY 1363
Db 1416 EAAALSRV--GQGSRLDGMVKRQFHGLLVKRFHCARRNSKALPSQILLPAFVCVANT 1474
QY 1364 FSLIVPFGKYPLELOPMWNYNDYV-----EVSNDAPF-----DTGELLELNL 1408
Db 1475 VALSVIEIGDLPVLVSPSOYH--NTQOPGNFIPYANERREYRLRLSDASPOQLVTF 1533
QY 1409 TKDPGEGTROM-----EGNPI----- 1424
Db 1534 RLPSGVAGATVLSKPSANGSLGPTLNLSCGESRLARFPDSMCLSEFTQGLPLSNFVPP 1593
QY 1425 -----PD-----TPQAGGEEMWTAP--VQOTIMDLFONONMNMNONSPAC 1463
Db 1594 PSPSPSDSPSPDEDIQAMNVSLPPTAGPBMNTSAPSLPRLVREYR-----C- 1641

```

```

QY 1464 QCSSDKTKKMLPYCPPRAGGLPPORRONTADILLODITGRNISDYLVKTYVOIIAKSLN 1523
Db 1642 TCSNOCYGSF--CPSSVYG-HPQMRVYTGDLITDITGHANSEYILFISDRF----- 1690
QY 1524 KIWNEFRYGFSLGVSNTOALPPSOEVNDALIKOMKHLKLAKDSSADRPLNSLGRFMTG 1583
Db 1691 ----RLHRYGATIFG--NVLSISIPASGTRAPMVRK----- 1721
QY 1584 LDRNNNVKVPFNKGMHAISSPLVNNALLRANLQKE--NPSHYGTAFNHPNLNLTQXO 1642
Db 1722 IAVRAAOVEYNNKGMHSMPTYLNSLNNALLRANLRKSKGNPAAYGITYNNHMTGAS 1781
QY 1643 LSEVALMTTSVDVLVSTICVIFAMSFVPASFPVFLIOERVSKAKHLOFISGVKPIVWLN 1702
Db 1782 LS-LDYILOGTDVIAIFILVAMSFPASFPVFLVLAKESTKAKHLOFVSGCNPIIYWLN 1840
QY 1703 FVWDMCNVVPATLVIIIFCPOOKSYVSTNLPLVALLLLYGMSTPLMTPASVFEXI 1762
Db 1841 YVMDLVNLVPAATCVIILFVFDLPATVSTPNPAVLSTFLYGWSITPIMVPSFWFEV 1900
QY 1763 PSTAYVVLTVSNLFIGINGSVATFVLELT--DKLNNINDILKSVFLPHPCLGGLD 1821
Db 1901 PSSAYVFLVILNFIIGLTAIVAFLLQLFEHDKDLKVNYSYLSCLFIPNVLGHLME 1960
QY 1822 MVKQAMADALERPGE--NRPVSLMDLVGRNLFAMAVEGFVFLITVLYRFFTRPRP 1880
Db 1961 MAYNEYINEYAKIGQFDKKSPPFEMDIYTRGLVAMAVEGVFLITIMQYIFLRPRQ 2020
QY 1881 VNAKLSPLNDEDEDREKORILLDGGGNDILEIKELTKYRKK--RKPADRICVGP 1937
Db 2021 MPVSTKVEED--DVVVAEBORVLEGDADNDMVKILENTKYKSKIKGRIILVADRLICLV 2079
QY 1938 PGECFGLGVNAGKSTFPMKLTGDTVTYTRGDAFLNKNSTLSNHEVHOMMGVCPQDAI 1997
Db 2080 PGECFGLGVNAGKSTFPMKLTGDESTTGEAVNHSVSKELLOQOOSLGTCPOCDAL 2139
QY 1998 TELLTGREHEFFALLRPEKEVYKGEVAMIRKILGVYKGEKAGVNSGNNRKLSTAM 2057
Db 2140 FDELTAREHLOLYRLKLGISWKDEARVAKMALEKLELTAKDPAGYSGNNRKLSTAI 2199
QY 2058 ALIGPPVVEIIDEPTTGMDPKARFIMNCALSVKBEBSVYVLSHSEBECALCTRATIM 2117
Db 2200 ALIGPAFIFLDEPTTGMDPKARFIMNLLIDLITKRSVVLVLSHSEBECALCTRATIM 2259
QY 2118 VNGRPRLGVSVOHLKKNRGDGYITVVRJAGSNPLKPYOFGFAGFVSLKEKRRMLQ 2177
Db 2260 VNGRLKLGISTQHLKKNRGDGYITVVR--TKSSQSVKDYVRFNRPFEALKEKRNHTKVQ 2318
QY 2178 YOLPSSLSLARIFSILSQRKRLHIEDYSVQTTLDQVFNFAKQSDSD 2227
Db 2319 YOLKSEHSLAQSVMKQVSGVLGIEDYSVQTTLDNVFNFAKQSDN 2368

```

RESULT 10

```

Q9UPU0 PRELIMINARY: PRT: 1529 AA.
ID Q9UPU0: Q9NSL2:
AC 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, last annotation update)
DE K1A1062 protein (Fragment).
GN K1A1062 OR DKFZP547p193.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=99397452; PubMed=10470851;
RA Kikuno R., Nagase T., Ishikawa K., Hirosewa M., Miyajima N.,
RA Tanaka A., Kotani H., Momura N., Ohara O.;

```

"Prediction of the coding sequences of unidentified human genes. XIV. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro".
 RL DNA Res. 6:197-205(1999).
 RM (2)
 RN SEQUENCE OF 1157-1529 FROM N.A.
 RC TISSUE-BRAIN:
 RA Bloecher H., Boecher M., Brandt P., Mewes H.W., Weil B., Wiemann S.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB028985; BAB3014.1; -
 DR EMBL; AL162060; CAB82398.1; -
 DR InterPro; IPR003439; ABC_transporter.
 DR InterPro; IPR000561; EGF-like.
 DR InterPro; IPR000566; Lipocalin_cyFABP.
 DR Pfam; PF00005; ABC_tran; 2.
 DR ProDom; PD00006; ABC_transporter; 2.
 DR PROSITE; PS00211; ABC_TRANSPORTER; UNKNOWN_1.
 DR PROSITE; PS00222; EGF_1; UNKNOWN_1.
 DR PROSITE; PS00213; LIPOCALIN; UNKNOWN_1.
 KW Hypothetical protein.
 FT NON_TER
 FT SEQUENCE 1 1529 AA; 170059 MW; BCED65F5E4AAC1A CRC64;
 Query Match 28.4%; Score 3345.5; DB 4; Length 1529;
 Best Local Similarity 46.7%; Pred. No. 7.8e-211;
 Matches 721; Conservative 207; Mismatches 381; Indels 235; Gaps 33;
 QY 836 GVMVYIAVPGGIGIRPWPYPCOTKSYWCF--ESDESHSGSNOKRMS-----ETC 887
 Db 1 GILMYIAVHPGMYGLPRPWFYLPQKSYWLSGRTTEWMSWMPWARTPRLSWEEDQAC 60
 QY 888 -----MEEPYHLKGVSIQNLVYKVRDGMKVAVDGIALNFEGQITSEFGLN 935
 Db 61 AMESRREETRGMEEPYHLPLVYCVDKLPYKYDKDKLNLKSLNLYENQVVSFEGHN 120
 QY 936 GAGGTTMSILTLGFPPTSGTAYILGKDISEMSTIQNLGVCPQHNVLPDMLTVEEHIW 995
 Db 121 GAGTGTSILTLGFPPTSGATYTGHDIREMDEIRKNGCMCPQHNVLPDLTVEEHLW 180
 QY 996 FYARLTKSEKHAEMEMALDVLPSKSLKSTQSLSGMOKLSVALAFVGSVVI 1055
 Db 181 FYSLKSMAGQELIRREMDKMTEDLEL-SNKRHSILVQLSGMKRKLISVALAFVGSVAIT 239
 QY 1056 LDEPTAGVDPYSRGIVELLIKYRQGTIILSTHMDADVLDGRILAIISGKLCCVSS 1115
 Db 240 LDEPTAGVDPYARRAIWDLILKYKPGRTIILSTHMDADVLDGRILAIISGKLCCGSP 299
 QY 1116 LFLKNQGTGYLLVLYKKDVSSLSGCRNSSSTVSYLAKKEDVSQSSDAGLCSHESDT 1175
 Db 300 LFLGTYGDDGIRLLVLRPAEPG-----GPQEPGLASSPPGRA 337
 QY 1176 LTIQVDS--AISNLRKHVSARLVEDIGHELTIVLYPEAKEGAFVLEFIEDRLSDIG 1233
 Db 338 PLSSCELOYSQFIRKHVASCLIVSDTSTELSTILPEAAKKAFELPQHLERSLDALH 397
 QY 1234 ISSIGISETILEELFLKVAEE-----SGVDA-ETSDGTLTP----- 1267
 Db 398 LSSFGIMDTLLEEVFLKVSSEDSLENSADYKESRKDVLPGAEGAPSGCHGANLARC 457
 QY 1268 -----ARRNRRA-FGDRQSGCLRPFTEDDAADPNSSDIDPRESREIDLS 1309
 Db 458 ELTQSQASLOSASSVSGARDEGAGTYDYGDRPLF-DMPQDPD--NNSLQVEVEAALS 514
 QY 1310 GMDQGSYGVKMKLTQOQFVALLMKRLLTARRSRKGFQAIYLPVAFVCIATVLSLVP 1369
 Db 515 RV-GQGRKIDGWLKRVQFHLVVRKFCARRNSKALFSQILLPFFVCVAMATVALSP 573
 QY 1370 PFGKYPSLELQPMKMYNQYT-----FVSNDAPE-----DTGLELLNALITDQPF 1414
 Db 574 ETIDLPPLVLPISOYH-NYTPQRGNEFLPYANBERREYRILSPDASPOQVAFRILPSCV 632
 QY 1415 GTRCM-----EGNPI----- 1424

Db 633 GATCVLKSPPANGSLGPTLLNLSGSESRLLAARFDSKLESTQGLPLSNFVPPSPAPS 692
 QY 1425 -----PD-----TPCAGEEWTAP-VPQITMDLFQNGNMTMQSPACCCSDK 1469
 Db 693 DSPASPDDEDLQANVSLPPTAGEMWTSAPSLPRLVREVR-----CTCSAG 740
 QY 1470 IKKMLPVCPPGAGGLPPQPKONTADILDLGRNISDYLVKTYVOITAKSLKRIWNE 1529
 Db 741 TGFS-----CPSSVVG-HPPMRVYTGILDTLGHNVSEYLLFTSDRF-----RL 785
 QY 1530 FRYGCSGLVSNTOALPPOSEVDAIKOMKHLKLAKDSADRFLNSLGRFMTGLDTRNN 1589
 Db 786 HRRGATTFG--NVLKSTIPASFGTRAPMYRK-----TAVRBA 820
 QY 1590 VKVFNKNGHAISSFLNINNALIRANLQGE-NPSHYGITAFAHPLNLTQKLSFVAL 1648
 Db 821 AQVEYNNKGHSMPTVLSNNAIIRANLPKSKGNPAAGITVTHNPMKTSASLS-LDY 879
 QY 1649 MTSYDVIVSICVIFAMSVVPASFVYFLQERYSKAKHLQFISGVKPYIYWLSNFVDMC 1708
 Db 880 LLOGTVVTAIFLIYAMSFVPASFVYFLVAKSTRAKHLQFVSGCNPIIYWLANYWDM 939
 QY 1709 NYVPATLVIIFIFICPOQSVYSTNLPLVALILLVGMSTPLWYPAFVKIPSTAYV 1768
 Db 940 NYLVPAITCCVILIFVDPDPAITSPTNFAVLSFLILGMSITPIMYPASFWEVSSAYV 999
 QY 1769 VLTSVVLFTINGSVATFVLELT-DNKLININDILKSVELLPFHCGLIDVKNQA 1827
 Db 1000 FLVIVNLFGLITATVATFLLQFEHDKDLKLVNYSKSCFLIEFPYNGHGHMEVANEY 1059
 QY 1828 MADALERGE-NRFVSPISPLDVLGRNLFMAVEGVVEFLITVLYQREFIRRPYNALTS 1886
 Db 1060 INEYAKIQOPDKMSPEFMDIVTQGLVAMAAGEVGVGLITMCOYNFLRRQRPYSTK 1119
 QY 1887 PLNDEDEDVRERORILLDGGQNDILEIKELTKIYRRK---RKPAVDRIQVIGPGEFCG 1943
 Db 1120 PVED-DVDASEERORVLRDADNDNWKIENLTKVYKSKRIGILLAVDLCLGVRGEFCG 1178
 QY 1944 LIGVNGAGKSTFEMKLTGTTVTRGDAPLKNKLSISNHEVQNNKGYCPQDQATITELLT 2003
 Db 1179 LIGVNGAGKSTFEMKLTGTTVTRGDAPLKNKLSISNHEVQNNKGYCPQDQATITELLT 1238
 QY 2004 REHVEFFALLRGVPERKEVKGEMAIRKILGYKGEKAYAGNVSGNKKRLSTAMALGGP 2063
 Db 1239 RHLQIYTLRLRISMKDEARVYKALKELETKYADKAGTYSGGNKKRLSTALALIGTP 1298
 QY 2064 PVEFLDEPTTGMDPKARRFLMNCALSVYKEGKSVYVLTSHSMECEALCTRNAIMVNGFR 2123
 Db 1299 AFIFLDEPTTGMDPKARRFLMNLIDILIKTGRSVYVLTSHSMECEALCTRLAIIVNGRLR 1358
 QY 2124 CLGSVOHLKNRFGDGYTVYRIAGSNPDLKVPQDFGLAFTGSYLKEKHNMLDYOUPSS 2183
 Db 1359 CLGSIQHLKNRFGDGYTMIVR-TKSSQSVKDVYRFNRPFEMALKENHHRKYVOYLKSE 1417
 QY 2184 LSLAIRFSILSSKKRLHIEDYSQTTLDQVFNFAKDDSD 2227
 Db 1418 HTSLAQVFSKMQVSGVIGIEDYSQTTLDQVFNFAKSDN 1461
 RESULT 11
 096J73
 ID 096J73 PRELIMINARY; PRT: 2277 AA.
 AC 096J73;
 DT 01-DEC-2001 (Tremblrel. 19, Created)
 DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE ATP-binding cassette transporter family A member 12.
 GN ABCA12..
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxId=9606;
 RN [1]

[illegible]

Db	869	FFPPIVATVENEISLYVKLFMSLISPTAFSAOQ IAREEEOGIGLOMENMYSPODD	928
Qy	818	GFNLITSIS-----MMLFDFLYCVMTWYIEANFPGQYGIIRWMPRCIKSYW---FG--E	868
Db	929	-----TTSFGMLCCLLIADSFIFYFLAMVBNVPFGYGAARWEPYLLPSTWKEFGCAE	984
Qy	869	ESDEKS-----HFGSNOKRMEICIMEEPHTLKJGVSIOMLVVYRGMCA	915
Db	985	VKPEKSNGLMFTNMQMNTNPASAPREPMSSNIEPRPKLTGVALHGVTKIY--GSKVA	1042
Qy	916	VDGLALNFEQOITSFLGHNGACKTTTMSITGLPEPPISTGAYILGKDIRSEMSTIRONL	975
Db	1043	VDNLNLFYEGHITSILSGPAGACKTTTISMLTGLFASAGTFYVYCKDIKTDHLTVKRM	11022
Qy	976	GVCSOHNVLFMDLTVEEHIMFYARK--GISEKHVKAEMOALUDVYPPSKSLKSTSQL	10333
Db	1103	GVCSOHNVLSYTLTKKHLHLLYSIGVPMWTKOLHEEVKRLKDTGLXSHRK--RVSTL	1151
Qy	1034	SGCMOKLSVALAFVSGSKVLIIDEPFAGVDVSRKGIWELLKRYROGRTIILSTHNDE	10933
Db	1162	SGMKRKRTISIALIGCSRVYIIDEPSTGDPVCSRRSINDVISKTKARTIILSTHHDE	12211
Qy	1094	ADYIGDRAIISHRKLCQVSSSLFLKNOIGTGYYLTLYKKVDSSLSGCRNSSSVSYLK	11533
Db	1222	AEVSDIRAFLEOGGLMCCSPPELKEARDGQHLLTKK-----	1261
Qy	1154	KEDVSOSSDAGCSGHESDTLTIDVSAISNLIRKXVEARLVEDIGHELTYVL--PYEA	12121
Db	1262	-----KSPNLNANAACDYM-----AVTAMIOGHLPEALKEDIGELTYVLPST	1307
Qy	1213	AKEGAFVELPHEIDRLSDIGISSYGISETTEELFLPKVAEESGYDAETSDGTLPARR--	1270
Db	1308	KVSAYSLIRALDNGMDLNICGYISTDYEEVFLNLTKESSORNSAMSLHEHLKOKIG	1367
Qy	1271	NRRAFQKQSCLPFTEDDADPNDSOIPRESEDTLKGMDGKSQYOVKQKGLKQOOVF	1330
Db	1368	NSMANGIS-----TYDDLSVSSSNTD---KDXKILF-----RGE-RUDFGILLKIKM	14122
Qy	1331	ALLKRLDIARSKRKGFFPAQIVLPAVEFCIALVESLIVPEFKRPSLELOPMY--NBQY	1388
Db	1413	AILIKRPHHTBRNKKGLIAQVILPIVEVTAMGLGTRNNSSYREIOLSPSLYCTSEOT	14722
Qy	1389	TFVSNDAPEDTGIELLNALTCKDPRGTRKMEBNPLPTDPC--QAGEEMPTAARPOIM	1446
Db	1473	AFYAYMR-----STALYASAMWDEPGIDNMCLNLS---DLOCNKSLEWMTNSGEPITNF	1528
Qy	1447	DLFGNGNTMOPSPACQSSDKIKKMLPYCPGAGGLPPROKONTAADIIODLGRNIS	1506
Db	1527	GV-----CSCSEN-----VQCP--KKNYSPPHRTYSSQVYINLGOVHE	1565
Qy	1507	DLYKTYVOIIIAKSLKMKINWNEF--RYGGSFLGVSNTQALPPSOEVNDAIKOMKHLK	1563
Db	1566	NYLIST-----ANEFYOKRYOGMSFG-----LP	1588
Qy	1564	LAKOSSADRFNLISGRMTGLD--TRNNKQWPNFKMGMAISSFLYNINNALIRANLOGE	16222
Db	1589	LTKDIRD-----ITGVPAKRLTAKWYDPBGYHSLPALNSLNNFLLRVMKSKYD	16333
Qy	1623	NPSHYGITAFAHNPILNTRKQOLSEVALMTTSDVLSIYCIAFANSPASFYVFLIOERVS	1683
Db	1640	AARH-GIIMYSHAPRGVODD--EQATISSILDLIALSILMGSYSTTASFYVYVREHOT	1696
Qy	1683	KAKHLOPISGVKPYIYMLSNFVDMCMQVYVAPATLVIIIFCPOOKSYVSTNLPVAILL	1742
Db	1697	KAKOLOHISGAGVTCWYTNFETIDMVTYLDVAFSIGIIAIFKLPFAEYBNNLGAVSLL	1755
Qy	1743	LLYGSITPLMYPASFEKPIRSTAYVVLTSVNLFIGINSVATVLELFTDKLNN---	1799
Db	1757	LLFGYATPSMWYLLAGFHEHGAFTIYVCVNLFEGINSITSLSVYFLSKERPNPTLE	1816
Qy	1799	-INDLKSVFLIFPHEGLKRLIDMKVNOAMADALEFEGNFRVSPSLMDYIGRNLFMA	185
Db	1817	LSEFLKFIPLIFPOCGGVGLIELSOOGSVDOFLKAYGVENFETEMNKIGAMFVALY	1876


```

Db 1156 PSLLLVFEAFNVHAFTRDGHMADLLPLMLYGAIIPLMLYMSFFSAASTAYTRLLT 1215
QY 1773 VMLFGLINSVATFV---LEFTDKKINNINDLSVFLPHEPGLGLDMVKN----- 1825
Db 1216 FILLSGI---ATFIVTMTCPAVKLEELSKRTLDHVFLLVLDHGLGMAVSFEYENETR 1271
QY 1826 -----QAMADALERGENREVSPLSWDL--VGRNLFMAAVGVFEFLITVLQ----- 1871
Db 1272 RCTSESLAHKCKKNIOYOSFYAMSTPGYKVTSAAGGIIYLLPLFIENILMR 1331
QY 1872 YREFI---RPRVNAKL---SPLNDEDEDVRRERORILDGGQNDI---LEIKELTKIYR 1922
Db 1332 LRFELICAFRRRTTALVLOMRTSVLPEDDVAEERSKRLIPSDSMIDPLLINEISKYV- 1390
QY 1923 RRRKP---AANDRICVGPPECFGILGVNGAKSSPFKMLTGTJYTRGAPLANKSIISN 1980
Db 1391 DOAPPLADYRISLAVQKCEFGLLSNAGKTTFFKMLTGETTYSGAFAVGYSISSD 1450
QY 1981 IREHONMGVCPQFDAITLTLTGREHEFFALLRGVPEKEKVGEMALRLGLVKYGER 2040
Db 1451 IGVNRGMGCPQFDALDHMTGREMLVMYARLGRIPERLINAVCENTLRGLLEPHANK 1510
QY 2041 YAGNYSGKRRKRLSTAMALIGRPVFLDEPTTGMDPKARRFLMNCALSVKEGSAVLT 2100
Db 1511 LVKTYSGKRRKRLSTIGLIGEPVIFLDPSTGMDPVARRLLMDVARRESGKAIVIT 1570
QY 2101 SHSMECEALCTFMAIMVNGRFRCLGSVOHLKRGFGDYTI---VVRIASNDLKPVOE 2158
Db 1571 SHSMECEALCTRLALIMVGOFRCLGSPHLSKFSGSILOAKVRSKQALDEFKAF 1630
QY 2159 FGLAFPGVSLKEKRRNMLYOULPSSLSLARIFSLISQSKRLIEDYSVSGOTLDOVEV 2218
Db 1631 VDLTFPGSLIEDHQMVMVHILPGCDLSWAKVGLLEKAKKGVYDYSVQSLSIOVFL 1690
QY 2219 NFA 2221
Db 1691 SFA 1693

```

RESULT 13

```

QYVRC4 PRELIMINARY; PRT: 1713 AA.
ID QYVRC4;
AC QYVRC4;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE CG1718 protein.
GN CG1718.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OC NCBI_TaxID=7227;
RN 1111
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Milos G.L.G.,
RA Arif J.F., Agayani A., An H.-J., Andrews-Pfankuch C., Baldwin D.,
RA Ballig R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borcha D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,

```

```

RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Houston D., Houston K.A., Howland T.J., Mei M.-H., Ibeagha C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kenison J.A., Ketchum K.A.,
RA Kisko B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lamm P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclebo J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reibert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svrtkars R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weisenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
DR EMBL: AE003568; AF50837.1;-.
DR Flybase: FBgn0031170; CG1718.
DR InterPro: IPR003439; ABC_transportr.
DR Pfam: PF00005; ABC_tran. 2.
DR Prosite: PD00006; ABC_transportr. 2.
DR PROSITE: PS00211; ABC_TRANSPORTER; 1.
KW ATP-binding; Transport.
SQ
SEQUENCE 1713 AA; 192888 MW; 9DE2003BFB9DC1CA CRC64;
Query Match 21.1%; Score 2485.5; DB 5; Length 1713;
Best Local Similarity 32.3%; Pred. No. 4.1e-154;
Matches 622; Conservative 354; Mismatches 603; Indels 345; Gaps 49;

```

```

Db      529  M V V K G S I M N M E D E T T V L L G H N G A G K T T T I S M L G M F P T S G T A I L N G S D I T N I E G A R M 588
Qy      974  N L G V C P O H N V L F E D M L T V E E H I M F A R K L G L S E H V A K E M O A M L D V G P S S L K K T S O L 1033
Db      589  S L G I C Q H N V L D E M S V N H I R F S K M K L R G K A V D E A K L K M I E L - E D K A N A V S S K L 647
Qy      1034  S G G M O R K L S V A L A F V G G S V V I L D E P T A G V D P Y S R R G I W E L L K Y R O G R T I L S T H H D E 1093
Db      648  S G G M K R K L S C A L C G D T V V L C D E P S S G M D S A R R O L M D L I O Q R K V G R T L L T T H F M D E 707
Qy      1094  A V L G R I A I I S H K L C V G S S L F L K N O L G T G Y I L T V K K D V E S S L S G R N S S V A T L K 1153
Db      708  A V L G R I A I M C D E L K C O G T S F L K K O Y G S G R L Y S G V N L F Y R C T Y K T C D S L K I C V K 767
Qy      1154  K E D V S O S S D A G L S D H E S D T L I D V A S I N L I R K H V S E A R L V E D I G H E L T Y V L P E A A 1213
Db      768  R D D - - - - - C E T N E V A L I N K Y I P G L K P E C D I G A E L S Y O L P D S A S 806
Qy      1214  K E G A F E L F H E I D R L S D I G S Y G S E T T L E I F L K A E S G V A E T S O G T L P A R R N R 1273
Db      807  A K - - F E E M F G L E O S D E L H I N G Y G V I T S M E V E M K V A E - - - - - K D N T G 850
Qy      1274  A F G D K O S C L R - - - P T E D D A A D P N D S I D P E S R E T D L S G M O G S Y O V K G M K L T Q O Q F V 1330
Db      851  N I K D O H E I N G S G F R G E D - - - - - D N E S V O S D - - - G I F S E K R R L L Q I O L L S N C M K 898
Qy      1331  A L L M K R L I A R S R K G F A Q I V L P A V E C I A L V F S L I P P F G K Y S L E L O P M W N E O Y T F 1390
Db      899  A M L K K F L T W M N K L L I L I O N I M P V E F V V I T - - L I K T O G T F G - E L K P I - - - - - 945
Qy      1391  V S N D A P E D T G L E L L N A L I K D P G F T R C M E G N P I D T I P C O A E E W T T A P V P Q I T M D L - - 1448
Db      946  - - - - - T I S L - - - - - T O Y P L A V I V L D R S N 963
Qy      1449  F O N G M W M O N P S P A C O C S S D K I K M L P V C P P G A G S L P P O R K O N A D I L I O L T G N I S D Y 1508
Db      964  V O N G - - - - - T G - - - - - Y 970
Qy      1509  L V K T Y V O I I A K S L K K I W N E R Y G - G F S L G V S N T O A L P P S O E V N D A I K O K M K L K L A K D 1567
Db      971  E I A N K Y E D I A R S - - - - - Y G S N G L E L T G T O G F - - - E D Y I L D I G K T I O V R I N 1013
Qy      1568  S S A D R F L N S L G F M T G L D T R N N K V W F N K G M H A I S S F L A N V I N N A L I R A N I O K E N S H Y 1627
Db      1014  S - - R Y L - - - - - V A A T T E S K I T A M L N O A L H T P L I V N A V H N A I - - A D L F G - - - S S V 1059
Qy      1628  G I T T A N H P I N L T K O O L S E V A L M T S I V D V A S I C V I F A M S F V A S F V F L I O E R V S K A H L 1687
Db      1060  K I O V T N A P L P Y T T S I L S Q L S T G N L G T O L A S N L C F C M C F V S S I Y I I F L I K E R S R A K L L 1119
Qy      1688  O F I S G V K P V I W L S F W M D M C N Y V P A T L V I I I F C O O K S Y S S T N L P A L L L L X G W 1747
Db      1120  O P V G V K W W T M L S O F I C D P A S Y I V A L I V I T T V C F O E T S L S T F G E L G R Y L L L L F G 1179
Qy      1748  S T T P L M P A S E V F K I P S T A Y V A L T S V N L F I G I N S V A F V L - - E L F T D K N I N I N D I L K S 1805
Db      1180  A V L P I T Y I M S L F E F R E P A T G P A R V S I V N F C G M A L F I V V W S S E L F - D T K - - D I A D I L G W 1256
Qy      1806  V F L I P H C L G R G L D M Y K K N O A M A D A L E R F E G E N R V S P L - - - - - S M D L V 1849
Db      1237  I F R I P H S I L A M S I N K Y T N A T R A C A K A G A - - L P I L C E L V P O C N L K P F A M E E P 1293
Qy      1850  G - - R N L F M A V E G V F L I T V L I O Y R - - - - - F F I R - - - - - R P V N A K S P L I N D E D E D 1894
Db      1294  G V L P E T Y V M A A T G V F F I I I T V L E R L I N E I M F K I R O L I S K P P P I E G O L - - - - - D D D 1347
Qy      1895  V R R E F O R I L D G G O N D I - - - - - L E I K E L T K I Y R R K R K P A V D R I C V I P P G E C F G L L G V N G A 1950
Db      1348  V A N E R E R I L Q - M S S N E L A T K M L V I D R V T K Y Y G O F M - - A V N G V S L C V O E V E G F G L L G V N G A 1404
Qy      1951  G K S S T F K M L T D D T Y T R D A P L A N K N S I L S N I H E V H O M K G C P O P D A I T E L L T G E H A E V F 2010

```

```

Db      1405  G K T T T F K M T G D E R I S S G A A V O G L S L E S M N N S I Y K M I G Y C P O F D A L D L D P G R E V L R I F 1464
Qy      2011  A L L G V P E K E V K G E A I R K L G L Y K G E Y K A C Y N S G G N R K R I S T A M A L I G G P P V F L D E 2070
Db      1465  C M L G V O S R I R O L S E D L A S F G M K H I D K O T A Y S G N R K R I S T A I A V I G S P S V I L D E 1524
Qy      2071  P T G M D P K A R R F L M C A L S Y V K E G R S V L T S H S M E E C A L C T M A I M V N G R F C I S G V O H 2130
Db      1525  P T G M D P A A R Q L M N V C R I R D S G K S I V L S H S M E E C A L C T R A I M V N G E F C I S T O H 1584
Qy      2131  L K N F G G G T I V - - - - - R I A G S - - - N P - - - - - D L K V O D F P L A F 2163
Db      1585  L K N F S G L L I K I K V R N L E A L R Q A R L S G Y A R P D E Q T V P A O M S O R D I D A V E F E T E Y 1644
Qy      2164  P G S V L K E H R N M L O Y O P S S I S L A R I F S I L S O K R L H I E D V S V O T T L D O Y F V N F A R D 2223
Db      1645  P N S I L O E Y O G I L T F I P L T G V K M S R I F G L M E S N R O L N E D I S V S O T T L E I L E F A K Y 1704
Qy      2224  Q S D D 2227
Db      1705  Q R E D 1708

RESULT 14
ID      Q8W010      PRELIMINARY;      PRT;      1884 AA.
AC      Q8W010;
DT      01-MAR-2002 (Tremblrel. 20, Created)
DT      01-MAR-2002 (Tremblrel. 20, Last sequence update)
DT      01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE      ATP-binding cassette transporter ABCA1.
OS      Arabidopsis thaliana (Mouse-ear cress).
OC      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC      Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC      eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX      NCBI_TaxID=3702;
RN      [1]
RP      SEQUENCE FROM N.A.
RA      Forestier C.;
RT      "AtABCA1, a full size ABC transporter homolog of the mammalian
RT      cholesterol efflux regulatory protein ABCA1."
RL      Submitted (Apr-2001) to the EMBL/Genbank/DBJ databases.
DR      EMBL: AY032590; AK339643.1;
DR      InterPro: IPR003593; AAA_ATPase.
DR      InterPro: IPR003439; ABC_transporter.
DR      Pfam: PF00005; ABC_tran. 2.
DR      ProDom: PD000006; ABC_transportr; 2.
DR      SMART: SM00382; AAA; 2.
DR      PROSITE: PS00211; ABC_TRANSPORTER; UNKNOWN_2.
KW      ATP-binding.
SQ
SEQUENCE 1884 AA: 209477 MW: 93CA7AB4465D588 CRC64;

Query Match      17.7%; Score 2092; DB 10; Length 1884;
Best local Similarity 28.8%; Pred. No. 3,9e-128;
Matches 592; Conservative 307; Mismatches 601; Indels 556; Gaps 58;

Qy      489  N E T N O A I R T S - R F M C V D N K L - - E P I A T E V L W I N K S M E L L D E R K F V A G I V T G I T G S 545
Db      59  D E T N M M I D L S I K F P E L R L V T K I F K D I L E T Y I T S A H G V C S E V R N S N P K I K A V V F H 118
Qy      546  I E L P H V H V K Y K I R M - - - - - D I D N V E R T N K I K G Y W D P G P R A D P F - - - - - D M R Y 588
Db      119  E G G P H L F D Y S I R L N H T W A A G F P N V K S I M D T N - - - - - G P Y I N D L E M G I N T I P T M O Y 169
Qy      589  V M G F A Y I O D V Y E Q A L I R V - - - - - L N G T E R K T G V - - - - - Y M Q M P Y P C - - 626
Db      170  S F S G F L T L Q O V V D S F I F A S Q O N N D P L S H S M L S A L R E L P W L F S P S V I R M V F P T P R E 229
Qy      627  Y V D D I L R A W S M S M P E M L A M I Y S A V I I K I Y E K E A R L K E T R I N G D L N S I I M S F W 686
Db      230  Y T D E F O S I V K S V M G L L Y L L G F L P I S R L I S Y S V E K E Q K I R E G L Y M M G L K D E I F H L S W 289
Qy      687  I S S I L P L V S A G L A V V I I K L G M L P S D P S V V F E L S V A V V T I I O C F L I S T L F S R A M L A 746

```


RA lightning J., Lloyd C., McMurtry A., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Ritten L., Roopra A., Saunders D., Showkeen R.,
 RA Smaildon N., Smith A., Sonhammer E., Staden R., Sulston J.,
 RA Thierly-Mieg J., Thomas K., Vaudin M., Vaughan R., Waterston R.,
 RA Watson A., Weinstock L., Wilkinson-Sprat J., Wohlman P.,
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans";
 RL Nature 368:32-38(1994).
 RN (2)
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2.
 RA Becker M., Graves T., Yeakum M.;
 RT "The sequence of C. elegans cosmid Y39D8C."
 RL Submitted (Oct-1998) to the EMBL/Genbank/DBJ databases.
 RN (3)
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Waterston R.;
 RL Submitted (MAR-1999) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AF101313; AAC69223.1; .
 DR InterPro: IPR003439; ABC_transporter.
 DR Pfam: PF00005; ABC_tran; 2.
 DR ProDom: PD00006; ABC_transport; 2.
 DR PROSITE: PS00211; ABC_TRANSPORTER; UNKNOMN.2.
 SQ SEQUENCE 1802 AA; 20325 MM; 933DBAEEF1AA046 CRC64;

Query Match 17.5%; Score 2060; DB 5; Length 1802;
 Best Local Similarity 27.9%; Pred. No. 4,6e-126;
 Matches 551; Conservative 338; Mismatches 630; Indels 454; Gaps 52;

QY 385 YTPDPATQVAEVAENKTPQELAVFHDLGEMWELSPKWTGEMNSOEDVLRMLDSRD 444
 Db 119 YAPMTATQIMDKIONRY-----TANDL-----LNP-VAATIKGLNITAVPVNTMT 167
 QY 445 NHFPEQDGLDWTADIVAFNAHPEDVSSNGSVYTRPAPNETNOAIRTTSREMEC 504
 Db 168 YKGF-----TTEGEMVSMOGQFQS-----EC 189
 QY 505 VNINKEPIATEVWLINKSMELDERKFWAGIVTGITPGSELPHHKKYKTRMDIOWE 564
 Db 190 DN-----PLL-----AGIVPDDSLAKDLKNDKDFYITRLSNTH 225
 QY 565 KINKIKDGYWDGPRADPEEDMRVWG-----GRAY 595
 Db 226 RRSR-----NARGDSYPMDDTSVFAVQYSGPINDDNGSGFQWEGFMT 273
 QY 596 LODVVEQAIIRVLGTGEEK-----TGVYMOQMPYPCYVDITFLRWKSMPLMTLANITY 650
 Db 274 VORAVDAVATEITIGEDAOPLPLDSYOVSRFPFGYSTKI-IEIGAFEMFVIVFESMT 332
 QY 651 SYAVILKIGIVEKEARKETMRIMGIDNSILWFSWIFSLIPPLVASGLLVITIKGLLL 710
 Db 333 SYIIVRAVAVVEKEDRLKRYMKVMSLQFINVAHFTINAKTLFAVAVLIIILHFAVL 391
 QY 711 PYSDPSVVFELSVAVVTLTLOCFILSLFESRANLAACGGITVFTLVL----- 759
 Db 392 -KSDMTLMEVFLIMAYFAFDVYFAFMSFMSATSATLISVFWMLLYEWAFESSIDQT 450
 QY 760 -PYVLCVAMQDVGFTLITFASLSP-VAFGGCGYFALFEEQIGGVOMDLFESPVED 817
 Db 451 NPYPL-----GTRL-----INCLNPDIALLNGQLLAAYETQADGLMGWELTPPSDN 499
 QY 818 GNLNTLSISMFLDFTLVGMVTWTEAVFGQYGP-RRWFPCPKRSWF--GEESDEK 873
 Db 500 NTFGHALLALIVDGIIMILLTWIEAVIPGEGVPOKPMWF-VLPSPWTFNSSGSKYIDS 558
 QY 874 SHPGSNOKRMASITCEEPETHLKLGVSIONLVKYV-----RDGMVAVDGL 919
 Db 559 SDQFOOIYADHVKLEKPEPTDILPTINNVNLTKYGTSEFFKFLDCKGKSGEKRAVSNL 618
 QY 920 ALNFEGQITSLFNGAGKTTMSLTGLPPTSGTAYIIGKDIRSMSTIRONLIGCP 979
 Db 619 NLKMTFGQCYVLLGHNGAGKSTTSMITGVASPPSGSAIVNDPIRISLPIKTRKMGICP 678

QY 960 QANVLFDMLTVEEHIMWEYARLKLS--EKHVKAEMQMALDVLSPSKLKSQTSQSGM 1037
 Db 679 QYNLTFCMTWMEHLEFFAKLERTWDPPEARRELIARLIDF-----KAPFMGALISGG 733
 QY 1038 QKLSVALAFVGGSKYVILDEPTAGVPSRGRIGWELLKRYROGRITIIISTHMDADV 1097
 Db 734 KRKLSTALALIGSEVYMLDEPTSGMDPARGHETWLTIOREKERRITLITHEMEADLL 793
 QY 1098 GDRITAIISHGKICCVSSSLFKNOIGTYVLTLYKKDVESLSLSCNRSSTVSYLKKDS 1157
 Db 794 GDRITAIHAGQLECCGSPMLKQYGDGYHLTY----- 827
 QY 1158 VSQSSDAGLGSDESPTLTIVSAISNLRKHVSEARLIEDIGHETLYVLPYEAKEGA 1217
 Db 828 -----YDTSPPDVSKTDTIREYIPEAHVSYIGQETVYL--SATNRP 871
 QY 1218 FVELFHEIDRLSDLGSSISGISTTLEELFLAVAE-----EGVDAETSQGLPAR 1269
 Db 872 FPKLFKELEDHQTCGITSFVSTTMEVFLVGHADERYNYENGENDISE----- 925
 QY 1270 RNRRAFGDKQSLRPFTEDDADPNDSIDIPESRETDLSGMDGKSQYQKGLTQOQF 1329
 Db 926 -----MEKD--DPILQD-----RAQRYTGFTLQMOHA 953
 QY 1330 VALMKRLRLARRSRKGFPAQIVLPAVEFCIALVSLVPPFGKYPSELQPMYNEQYT 1389
 Db 954 KAMFYKRAIFEFRRKWQFLPOLVFPVAYLVLMFTSOVLPSVKE----- 997
 QY 1390 FVSDAPEDPTGTELLNALIKDGFGRCEMGPIDPTPOAQEEEMTAPVQITMDLF 1449
 Db 998 -----QDPQITSLA-----PFSQTK-KAG-----HLVSD-- 1020
 QY 1450 QNGNMTQNSPACQCSSDKIKMLPYCPGAGLPPPOKQMTADILDTLGRNISDY 1509
 Db 1021 -SGNYVTL-----LG-----SQNLSSNV 1038
 QY 1510 VKTYVQIIAASLKNKIWNFRYGGSLGYSNQALPPSEVNDALIKMKHLKLAKDS 1569
 Db 1039 QGTYTO-----LGVTOF-----VVDITSNVEKEFTMDQTM 1069
 QY 1570 ADR-----FLNSLGRMTGLDTRNNYKVFNNKMGHAISFLVNNATLRANLQK 1620
 Db 1070 GSRFGLHVALGFVPSMFNST--VSVPILKTFEENNFGLYTPALATTFDSMILSKOKK 1127
 QY 1621 GENDSHTGTAENHPLNLTKQOLEVALMTTSDVLVSCVIFPMSFVPSVASEVYLIQR 1680
 Db 1128 -----QYSTFAVNHPLPSPSTQDTLKNMNSDGAAPFLAAGLIVFAVCAVAGYSQFLTER 1182
 QY 1681 VSKAKHLQFISGVKPYVYUWLSNFMQCNVYVATLVIIIFICFOOKSVSSGN-LPVL 1739
 Db 1183 KRSKHMQLLSGIRPMWFMTATITMDPAWFLIKILCFDAIFETJNTATTHDFGVALILT 1242
 QY 1740 LLLLLYGSITPLMYPASEFEKIPSTAYVVLVSNDLFGINGSVATFVELFTDNKLNLI 1799
 Db 1243 LSFLLYQWMTALPPTYMQEFEESAPRGPMVIMYHILITOMISSIAPRIISQTSISDAGVL 1302
 QY 1800 NDILKSFLIFPFCIGRGLIDMVKNQAMADALER-----GENREVS 1842
 Db 1303 WSIIFA--WLEPYNISQIATVTFQENENRIACKKIDCTIPMEKAVTACCGTASERLVD 1360
 QY 1843 PLMSDLYVGRN-----LEMAVGVVEFLITYLQYRF-----IRPPVNAKSLPND 1890
 Db 1361 NVLP--VGNKKGILVYVFLAVDGFYIWIWEMRENDQTKLALIRCKKADNPIMDITD 1418
 QY 1891 -----EDEDVRRR--ORILDGGQNDLLEIKELTKYIRKKRKPAVDRCVGP 1938
 Db 1419 TDKVDERVEDSDVIAEKSVYQRL--ANNKNTALVSNLNVKMYGNP--AVGVMENHVS 1474
 QY 1939 GECEGLLVNAGAKSSTFKMLTQDITVTGDAFLKNKSLNSINHEHOMAGCPOFDAIT 1998
 Db 1475 KDCFGLGVNAGAKSTFQMLTGENSEISSGDYAVNGMSVKNMWRAGANTGYCPOYDALI 1534

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 7, 2003, 08:45:29 ; Search time 24 Seconds
(Without alignments)
2771.884 Million cell updates/sec

Title: US-09-595-526C-2
Perfect score: 11797
Sequence: 1 MACWPQLRLMKNLTPRRR.....VDVAVLTSFLDEKESYV 2261

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :
1: Issued_Patents_AA.*
2: /cgn2_6/ptodata/1/1aa/5A.COMB.pep.*
3: /cgn2_6/ptodata/1/1aa/5B.COMB.pep.*
4: /cgn2_6/ptodata/1/1aa/6A.COMB.pep.*
5: /cgn2_6/ptodata/1/1aa/6B.COMB.pep.*
6: /cgn2_6/ptodata/1/1aa/6C.COMB.pep.*
7: /cgn2_6/ptodata/1/1aa/6D.COMB.pep.*
8: /cgn2_6/ptodata/1/1aa/6E.COMB.pep.*
9: /cgn2_6/ptodata/1/1aa/6F.COMB.pep.*
10: /cgn2_6/ptodata/1/1aa/6G.COMB.pep.*
11: /cgn2_6/ptodata/1/1aa/6H.COMB.pep.*
12: /cgn2_6/ptodata/1/1aa/6I.COMB.pep.*
13: /cgn2_6/ptodata/1/1aa/6J.COMB.pep.*
14: /cgn2_6/ptodata/1/1aa/6K.COMB.pep.*
15: /cgn2_6/ptodata/1/1aa/6L.COMB.pep.*
16: /cgn2_6/ptodata/1/1aa/6M.COMB.pep.*
17: /cgn2_6/ptodata/1/1aa/6N.COMB.pep.*
18: /cgn2_6/ptodata/1/1aa/6O.COMB.pep.*
19: /cgn2_6/ptodata/1/1aa/6P.COMB.pep.*
20: /cgn2_6/ptodata/1/1aa/6Q.COMB.pep.*
21: /cgn2_6/ptodata/1/1aa/6R.COMB.pep.*
22: /cgn2_6/ptodata/1/1aa/6S.COMB.pep.*
23: /cgn2_6/ptodata/1/1aa/6T.COMB.pep.*
24: /cgn2_6/ptodata/1/1aa/6U.COMB.pep.*
25: /cgn2_6/ptodata/1/1aa/6V.COMB.pep.*
26: /cgn2_6/ptodata/1/1aa/6W.COMB.pep.*
27: /cgn2_6/ptodata/1/1aa/6X.COMB.pep.*
28: /cgn2_6/ptodata/1/1aa/6Y.COMB.pep.*
29: /cgn2_6/ptodata/1/1aa/6Z.COMB.pep.*
30: /cgn2_6/ptodata/1/1aa/6A.COMB.pep.*
31: /cgn2_6/ptodata/1/1aa/6B.COMB.pep.*
32: /cgn2_6/ptodata/1/1aa/6C.COMB.pep.*
33: /cgn2_6/ptodata/1/1aa/6D.COMB.pep.*
34: /cgn2_6/ptodata/1/1aa/6E.COMB.pep.*
35: /cgn2_6/ptodata/1/1aa/6F.COMB.pep.*
36: /cgn2_6/ptodata/1/1aa/6G.COMB.pep.*
37: /cgn2_6/ptodata/1/1aa/6H.COMB.pep.*
38: /cgn2_6/ptodata/1/1aa/6I.COMB.pep.*
39: /cgn2_6/ptodata/1/1aa/6J.COMB.pep.*
40: /cgn2_6/ptodata/1/1aa/6K.COMB.pep.*
41: /cgn2_6/ptodata/1/1aa/6L.COMB.pep.*
42: /cgn2_6/ptodata/1/1aa/6M.COMB.pep.*
43: /cgn2_6/ptodata/1/1aa/6N.COMB.pep.*
44: /cgn2_6/ptodata/1/1aa/6O.COMB.pep.*
45: /cgn2_6/ptodata/1/1aa/6P.COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6931	58.8	1375	US-08-665-259-26	Sequence 26, Appl
2	6931	58.8	1375	US-08-762-500-26	Sequence 26, Appl
3	3129.5	26.5	1457	US-08-665-259-27	Sequence 27, Appl
4	3129.5	26.5	1457	US-08-665-259-27	Sequence 27, Appl
5	2642.5	22.4	1684	US-08-665-259-25	Sequence 25, Appl
6	2642.5	22.4	1684	US-08-665-259-25	Sequence 25, Appl
7	2642.5	22.4	1704	US-08-762-500-75	Sequence 75, Appl
8	340.5	2.9	1280	US-08-583-276-19	Sequence 19, Appl
9	339.5	2.9	1280	US-08-583-276-19	Sequence 19, Appl
10	337	2.9	1279	US-08-784-649A-2	Sequence 2, Appl
11	330.5	2.8	1280	US-08-752-447-2	Sequence 2, Appl
12	330.5	2.8	1280	US-08-752-447-2	Sequence 2, Appl
13	315.5	2.7	1275	US-09-134-001C-369	Sequence 369, Ap
14	311	2.6	1275	US-09-120-513-2	Sequence 2, Appl
15	311	2.6	1275	US-09-450-105-2	Sequence 2, Appl
16	309	2.6	1302	US-09-134-001C-3570	Sequence 3570, Ap
17	301.5	2.5	1302	US-08-232-537-2	Sequence 2, Appl
18	290	2.5	1337	US-09-134-001C-5550	Sequence 5550, Ap
19	288.5	2.4	1349	US-08-612-734B-2	Sequence 12, Appl
20	285.5	2.4	1349	US-08-612-734B-2	Sequence 12, Appl
21	284	2.4	1349	US-09-627-376-12	Sequence 3824, Ap
22	283	2.4	1408	US-08-612-521-2	Sequence 3824, Ap
23	281.5	2.4	1408	US-09-134-001C-3292	Sequence 3292, Ap
24	276	2.3	1307	US-09-134-001C-3539	Sequence 3539, Ap
25	275	2.3	1307	US-08-395-246C-2	Sequence 2, Appl
26	266.5	2.3	1357	US-09-134-001C-4891	Sequence 4891, Ap
27	261	2.2	1334	US-08-996-545-2	Sequence 2, Appl

28	261	2.2	1334	US-09-328-320-2	Sequence 2, Appl
29	250	2.1	376	US-08-997-080-89	Sequence 89, Appl
30	250	2.1	376	US-08-997-362-89	Sequence 89, Appl
31	250	2.1	376	US-08-873-970-89	Sequence 89, Appl
32	250	2.1	376	US-09-095-855-89	Sequence 89, Appl
33	250	2.1	376	US-09-324-542-89	Sequence 89, Appl
34	250	2.1	376	US-09-324-542-89	Sequence 89, Appl
35	244.5	2.1	341	US-09-205-426-89	Sequence 4042, Ap
36	242	2.1	215	US-09-134-001C-4042	Sequence 4042, Ap
37	242	2.1	215	US-09-305-984-18	Sequence 18, Appl
38	242	2.1	215	US-09-073-541A-18	Sequence 18, Appl
39	240	2.0	1528	US-08-463-092B-6	Sequence 6, Appl
40	240	2.0	1528	US-08-462-109A-6	Sequence 6, Appl
41	240	2.0	1528	US-08-460-907B-6	Sequence 6, Appl
42	240	2.0	1528	US-08-463-179A-6	Sequence 6, Appl
43	240	2.0	1528	US-08-461-384B-6	Sequence 6, Appl
44	237	2.0	224	US-09-305-984-72	Sequence 72, Appl
45	237	2.0	224	US-09-305-984-74	Sequence 74, Appl

ALIGNMENTS

RESULT 1
US-08-665-259-26
Sequence 26, Application US/08665259
Patent No. 6028173
GENERAL INFORMATION:
APPLICANT: Landes, Gregory M.
APPLICANT: Burn, Timothy C.
APPLICANT: Comoros, Timothy D.
APPLICANT: Dackowski, William R.
APPLICANT: Van Raay, Terence J.
APPLICANT: Klinger, Katherine W.
TITLE OF INVENTION: NOVEL HUMAN CHROMOSOME 16 GENES,
TITLE OF INVENTION: COMPOSITIONS, METHODS OF MAKING AND USING SAME
NUMBER OF SEQUENCES: 73
CORRESPONDENCE ADDRESS:
ADDRESSEE: GENZYME CORPORATION
STREET: One Mountain Road
CITY: Framingham
STATE: Massachusetts
COUNTRY: United States of America
ZIP: 01701
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/665,259
FILING DATE: 17-JUN-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Dugan, Deborah A.
REGISTRATION NUMBER: 37,315
REFERENCE/DOCKET NUMBER: IG5-9.1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (508) 872-8400
TELEFAX: (508) 872-5415
INFORMATION FOR SEQ ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 1375 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-665-259-26

Query Match: 58.8%; Score 6931; DB 3; Length 1375;
Best Local Similarity: 97.2%; Pred. No. 0;
Matches 1336; Conservative 20; Mismatches 19; Indels 0; Gaps 0;

```

QY 887 CMEEBTHLKLGVSIQNIYKVVYRDGMKVAVDGLAIFEGQITTSFLGHNGAGKTTMSIL 946
    |||||
Db 1 CMEEBTHLRLGVSIQNIYKVVYRDGMKVAVDGLAIFEGQITTSFLGHNGAGKTTMSIL 60
QY 947 TGLFPPTSTATITLKGDIRSEMSITRONLGYCPDHNVLFDMLVEEHIMFYARLKGISEK 1006
    |||||
Db 61 TGLFPPTSTATITLKGDIRSEMSITRONLGYCPDHNVLFDMLVEEHIMFYARLKGISEK 120
QY 1007 HYKAEEMQALDVGIPSSKLSKTSQLSGMOKRLSVLAFVGGSKVYLDEPTAGVPPY 1066
    |||||
Db 121 HYKAEEMQALDVGIPSPKLSKTSQLSGMOKRLSVLAFVGGSKVYLDEPTAGVPPY 180
QY 1067 SRRGIMELLKTRGRTIILSTHMDVADRIATISHKGLCCVSSLEFLKMOLGTY 1136
    |||||
Db 181 SRRGIMELLKTRGRTIILSTHMDVADRIATISHKGLCCVSSLEFLKMOLGTY 240
QY 1127 YTLTKKDVESLSSCRNSSSTVSYLKKEDYSOSSDAGISGSHESDTLIDVSAISNL 1166
    |||||
Db 241 YTLTKKDVESLSSCRNSSSTVSYLKKEDYSOSSDAGISGSHESDTLIDVSAISNL 300
QY 1187 IRKHVSEARLVEDIGHELTYLPEAKEGAFVELFHEIDRLSDLGISYISSETTLEE 1246
    |||||
Db 301 IRKHVSEARLVEDIGHELTYLPEAKEGAFVELFHEIDRLSDLGISYISSETTLEE 360
QY 1247 IFLKAESGVDAETSDGTLPARRRRKRAFQKQSLRFTEDDAADPVDSDIDPESRETD 1306
    |||||
Db 361 IFLKAESGVDAETSDGTLPARRRRKRAFQKQSLRFTEDDAADPVDSDIDPESRETD 420
QY 1307 ILSGMDGKGYOVKGMKLTQOQFVALLMKRLILARRSRKGFPAQIVLPAYVCAIATVSL 1366
    |||||
Db 421 ILSGMDGKGYOVKGMKLTQOQFVALLMKRLILARRSRKGFPAQIVLPAYVCAIATVSL 480
QY 1367 IVPFGKTPSELEQWMMNEQTYEVSNDAPEDTGLELNLALNLPDGFTRCMEGNPIPD 1426
    |||||
Db 481 IVPFGKTPSELEQWMMNEQTYEVSNDAPEDTGLELNLALNLPDGFTRCMEGNPIPD 540
QY 1427 TPQAGEEEMTAPVPOTIMDLFQNGNMTMONPSPACCCSSDKIKMLPYCPGAGGIPP 1486
    |||||
Db 541 TPQAGEEEMTAPVPOTIMDLFQNGNMTMONPSPACCCSSDKIKMLPYCPGAGGIPP 600
QY 1487 PQRKONTADILLODLTGRNISDYLVKTYOIIAKSLKKNIWNEFRYGGFSLGVSNTQALP 1546
    |||||
Db 601 PQRKONTADILLODLTGRNISDYLVKTYOIIAKSLKKNIWNEFRYGGFSLGVSNTQALP 660
QY 1547 PSEOENDAIKOMKRLKLAKDSSADRFNLISGREMTGIDTRNNKIVWPNKGMATISFL 1606
    |||||
Db 661 PSEOENDAIKOMKRLKLAKDSSADRFNLISGREMTGIDTRNNKIVWPNKGMATISFL 720
QY 1607 NVINNALIRANLQGENPSHYGTAFFNHPMLTKOOLSEVALMTTSVDVLSICVIFAMS 1666
    |||||
Db 721 NVINNALIRANLQGENPSHYGTAFFNHPMLTKOOLSEVALMTTSVDVLSICVIFAMS 780
QY 1667 FVPASVVELLOERYSKAKHLOFTISGVKPYTYMNSNVWDMCNVVAIVLIIIFTCFOQ 1726
    |||||
Db 781 FVPASVVELLOERYSKAKHLOFTISGVKPYTYMNSNVWDMCNVVAIVLIIIFTCFOQ 840
QY 1727 KSYVSTNLPVALILLLLYGSMTPLMPASFVFKISTAVVLTSVNLFINGISVATE 1786
    |||||
Db 841 KSYVSTNLPVALILLLLYGSMTPLMPASFVFKISTAVVLTSVNLFINGISVATE 900
QY 1787 VLEFLTDNKLNDILKSVFLIPHFCLGRLIDMYKKNQAMADALERFGBNRFVPLSW 1846
    |||||
Db 901 VLEFLTDNKLNDILKSVFLIPHFCLGRLIDMYKKNQAMADALERFGBNRFVPLSW 960
QY 1847 DLVGNLFAMAVEGVFLITVLIOYRFPPIRPPVNAKLSPLNDEDDVAREQRITDGG 1906
    |||||
Db 961 DLVGNLFAMAVEGVFLITVLIOYRFPPIRPPVNAKLSPLNDEDDVAREQRITDGG 1020
QY 1907 GONDILEIKELTKLYRRKRRPAVDRIQGIPEGECFGLGVNGAGSSPFKMLTGDTTV 1966
    |||||
Db 1021 GONDILEIKELTKLYRRKRRPAVDRIQGIPEGECFGLGVNGAGSSPFKMLTGDTTV 1080
QY 1967 RGDAPFLKNSILSIHVEHONMGYCPQFDATITELTGREHVEFPALLRGVPEKEVGKVE 2026

```

```

Db 1081 RGDAPFLKNSILSIHVEHONMGYCPQFDATITELTGREHVEFPALLRGVPEKEVGKVE 1140
QY 2027 MAIRKGLVYKGERKYNAGNSGKRRKLSMALIGSPVVFIDEPTTGMDPARREFLWNC 2086
    |||||
Db 1141 MAIRKGLVYKGERKYNAGNSGKRRKLSMALIGSPVVFIDEPTTGMDPARREFLWNC 1200
QY 2087 ALSVYKGRSVVLTISHMECEALCTRMALMVNGFRRCISGVQHLKRRGDCYTVVRIA 2146
    |||||
Db 1201 ALSVYKGRSVVLTISHMECEALCTRMALMVNGFRRCISGVQHLKRRGDCYTVVRIA 1260
QY 2147 GSNPDLKPVODFFGLAFPGSVLKEKRRNMLQYOLPSSLSTARISILSOSKRLHIEDY 2206
    |||||
Db 1261 GSNPDLKPVODFFGLAFPGSVLKEKRRNMLQYOLPSSLSTARISILSOSKRLHIEDY 1320
QY 2207 SVSOTITLDQYFVNFARQSDSDHLKDLSLKKNQYVDVAVLTSFLODEVKKESYV 2261
    |||||
Db 1321 SVSOTITLDQYFVNFARQSDSDHLKDLSLKKNQYVDVAVLTSFLODEVKKESYV 1375

```

RESULT 2

```

US-08-762-500-26
; Sequence 26, Application US/08762500
; Patent No. 6030806
; GENERAL INFORMATION:
; APPLICANT: Landes, Gregory M.
; APPLICANT: Burn, Timothy C.
; APPLICANT: Connors, Timothy D.
; APPLICANT: Mackowski, William R.
; APPLICANT: Van Raay, Terence J.
; APPLICANT: Klingner, Katherine W.
; TITLE OF INVENTION: NOVEL HUMAN CHROMOSOME 16 GENES,
; TITLE OF INVENTION: COMPOSITIONS, METHODS OF MAKING AND USING SAME
; NUMBER OF SEQUENCES: 83
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENZYME CORPORATION
; STREET: One Mountain Road
; CITY: Framingham
; STATE: Massachusetts
; COUNTRY: United States of America
; ZIP: 01701
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/762,500
; FILING DATE: 09-DEC-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/665,259
; FILING DATE: 17-JUN-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/10469
; FILING DATE: 17-JUN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Dugan, Deborah A.
; REGISTRATION NUMBER: 37,315
; REFERENCE/DOCKET NUMBER: 195-9.3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (508) 872-8400
; TELEFAX: (508) 872-5415
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1375 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-762-500-26

```

Query Match

58.8%; Score 6931; DB 3; Length 1375;

Best Local Similarity 97.2%; Pred. No. 0;
Matches 1336; Conservative 20; Mismatches 19; Indels 0; Gaps 0;

```

QY 887 CMEEPPTLLKGVSTONLVKYYRGMKAVVDGLALNFEQGITSLFLNGNGKTTTSTL 946
Db 1 CMEEPPTLLKGVSTONLVKYYRGMKAVVDGLALNFEQGITSLFLNGNGKTTTSTL 60
QY 947 TGLPPTSGTAVIIGKIDRSEKSTIRONLGYCPQHNVLFDMLTVEEHWFAARLKGLSEK 1006
Db 61 TGLPPTSGTAVIIGKIDRSEKSTIRONLGYCPQHNVLFDMLTVEEHWFAARLKGLSEK 120
QY 1007 HKKAMEOMALDVGLPSSKLSKTSQSLSGMQRLSVALAFVGSKVYIIDEPTAGVDPY 1066
Db 121 HKKAMEOMALDVGLPSSKLSKTSQSLSGMQRLSVALAFVGSKVYIIDEPTAGVDPY 180
QY 1067 SRRGIMWELLKRYROGRTIITLSTHHMDEADVIGDRITAIISHKLCVGSFLKNOGLTGY 1126
Db 181 SRRGIMWELLKRYROGRTIITLSTHHMDEADVIGDRITAIISHKLCVGSFLKNOGLTGY 240
QY 1127 YTLVKKVDESSLSGCRSSSTVSYLKKEDSVQSSSDAGSGSHESDTLTIDVSAISNL 1186
Db 241 YTLVKKVDESSLSGCRSSSTVSYLKKEDSVQSSSDAGSGSHESDTLTIDVSAISNL 300
QY 1187 IRKHVSEARLVEDIGHELTYVLPYPAKEGAFVELFHEIDRLSDLGISYGISSETTLEE 1246
Db 301 IRKHVSEARLVEDIGHELTYVLPYPAKEGAFVELFHEIDRLSDLGISYGISSETTLEE 360
QY 1247 IFLKAEESGVAFETSDGTIPARRRRAPGKOSCLRPETEDDADAPDSDIDPESRPTD 1306
Db 361 IFLKAEESGVAFETSDGTIPARRRRAPGKOSCLRPETEDDADAPDSDIDPESRPTD 420
QY 1307 LLSGMDGSGSYOVKMKLTQOQFVALLMKRLLIARRSRKGEFAQIVLPAVVCIALVESL 1366
Db 421 LLSGMDGSGSYOVKMKLTQOQFVALLMKRLLIARRSRKGEFAQIVLPAVVCIALVESL 480
QY 1367 IVPFGKYPSELQPMWYNEOYTFVSNDAPEDTGTELLMALTKDPGFGTCROMEGNPIPD 1426
Db 481 IVPFGKYPSELQPMWYNEOYTFVSNDAPEDTGTELLMALTKDPGFGTCROMEGNPIPD 540
QY 1427 TPCQAGEEEMTAPAPORTIMOLFONGNMNMONSPACCCSSDKIKKMLPVCPPGAGGLPP 1486
Db 541 TPCQAGEEEMTAPAPORTIMOLFONGNMNMONSPACCCSSDKIKKMLPVCPPGAGGLPP 600
QY 1487 PORKONTADIILODLTGRNISDYLVKTYVOIITAKSLKNKIWNEFRYGGFSLGVSNTQALP 1546
Db 601 PORKONTADIILODLTGRNISDYLVKTYVOIITAKSLKNKIWNEFRYGGFSLGVSNTQALP 660
QY 1547 PQOEYNDATIKOMKHLKLAKOSSADREPLNSIGREMTGLDTRNNVAVWFNNGKWAHAISSFL 1606
Db 661 PQOEYNDATIKOMKHLKLAKOSSADREPLNSIGREMTGLDTRNNVAVWFNNGKWAHAISSFL 720
QY 1607 NVINNAIILRANLOKGENPSHYGITAFNHPNLNTKOLSEVALMTSVIVIASICVIFAMS 1666
Db 721 NVINNAIILRANLOKGENPSHYGITAFNHPNLNTKOLSEVALMTSVIVIASICVIFAMS 780
QY 1667 FVPASVVFVLIQERYSKAKHLOFISGVKPVYIWLSNFVWDMCNVVPATLVITITCFQO 1726
Db 781 FVPASVVFVLIQERYSKAKHLOFISGVKPVYIWLSNFVWDMCNVVPATLVITITCFQO 840
QY 1727 KSYVSTNLPVIALLLLVGMSITPLMPASVVERKIPSTAYVVLTVSNLFGINGSVATF 1786
Db 841 KSYVSTNLPVIALLLLVGMSITPLMPASVVERKIPSTAYVVLTVSNLFGINGSVATF 900
QY 1787 VLELFDNKLNNINILIKSVFLIPHFCGLRGLIMVKNQAAADLFRGENRFPVSPISM 1846
Db 901 VLELFDNKLNNINILIKSVFLIPHFCGLRGLIMVKNQAAADLFRGENRFPVSPISM 960
QY 1847 DLVGRNLFMAAVEGVFFLITVLIQREFIRPRPVNAKLSPLNDEDEVDVRRRORILDOG 1906
Db 961 DLVGRNLFMAAVEGVFFLITVLIQREFIRPRPVNAKLSPLNDEDEVDVRRRORILDOG 1020
QY 1907 GONDIEIEIKELTKYRRKKRAVDRIICVIGPECEFGLLGVNAGAKSSTFKMLTJDTIVT 1966
Db 1907 GONDIEIEIKELTKYRRKKRAVDRIICVIGPECEFGLLGVNAGAKSSTFKMLTJDTIVT 11

```

```

Db 1021 GONDIEIEIKELTKYRRKKRAVDRIICVIGPECEFGLLGVNAGAKSSTFKMLTJDTIVT 1080
QY 1967 RGAFLNKNSTLINSIHVEHONMGYCPQDAITPELLGREHVEFPALLRGVPEKXGVGE 2026
Db 1081 RGAFLNKNSTLINSIHVEHONMGYCPQDAITPELLGREHVEFPALLRGVPEKXGVGE 1140
QY 2027 WAIIRKLGIVYKGEYAGN SGNKRLKSTAMALIGPPVFLDEPTTGMDPKARFLMNC 2086
Db 1141 WAIIRKLGIVYKGEYAGN SGNKRLKSTAMALIGPPVFLDEPTTGMDPKARFLMNC 1200
QY 2087 ALSVKEGRSVLTSMSMECEALCTRMALMVNGRRRCGLSVQHLKNRGDGTIVVRIA 2146
Db 1201 ALSVKEGRSVLTSMSMECEALCTRMALMVNGRRRCGLSVQHLKNRGDGTIVVRIA 1260
QY 2147 GSNPDLKPVODFFGLAFPGSVLKEKHRNMLOYPSSLASLAFISLSOSKRLHIEDY 2206
Db 1261 GSNPDLKPVODFFGLAFPGSVLKEKHRNMLOYPSSLASLAFISLSOSKRLHIEDY 1320
QY 2207 SVSQTTLDOYFVNFPAKDQSDDDLKDLSLHKNQTVYDVAVLTSFLODEKVESYV 2261
Db 1321 SVSQTTLDOYFVNFPAKDQSDDDLKDLSLHKNQTVYDVAVLTSFLODEKVESYV 1375

```

RESULT 3

```

US-08-665-259-27
: Sequence 27, Application US/08665259
: Patent No. 6028173
:
: GENERAL INFORMATION:
: APPLICANT: Landes, Gregory M.
: APPLICANT: Burn, Timothy C.
: APPLICANT: Connors, Timothy D.
: APPLICANT: Dackowski, William R.
: APPLICANT: Van Raay, Terence J.
: APPLICANT: Klinger, Katherine W.
: TITLE OF INVENTION: NOVEL HUMAN CHROMOSOME 16 GENES,
: TITLE OF INVENTION: COMPOSITIONS, METHODS OF MAKING AND USING SAME
: NUMBER OF SEQUENCES: 73
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: GENZYME CORPORATION
: STREET: One Mountain Road
: CITY: Framingham
: STATE: Massachusetts
: COUNTRY: United States of America
: ZIP: 01701
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/665,259
: FILING DATE: 17-JUN-1996
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Dugan, Deborah A.
: REGISTRATION NUMBER: 37,315
: REFERENCE/DOCKET NUMBER: IG5-9.1
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (508) 872-8400
: TELEFAX: (508) 872-5415
: INFORMATION FOR SEQ ID NO: 27:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1457 amino acids
: TYPE: amino acid
: STRANDEDNESS: not relevant
: TOPOLOGY: unknown
: MOLECULE TYPE: protein
:
: US-08-665-259-27

```

Query Match 26.5%; Score 3129.5; DB 3; Length 1457;
Best Local Similarity 46.0%; Pred. No. 3e-288;
Matches 676; Conservative 207; Mismatches 377; Indels 211; Gaps 28;

```

QY 888 MEEPTHLKGVSIGNLVKVRDGMKVAVDGLALNFEGQITSPFGHAGAKTTMTSILT 947
    ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
Db 1 MEEPTHLPIVVCYDKLTKVYKNDKKLALNKLNLLENYVAVSFLGNGACKTTMTSILT 60
QY 948 GLEPPTSGAVILGDIIDISESTIRONLGVCPQHNVLFDMLTVEHIMFWFARLKGSLSKH 1007
    ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
Db 61 GLEPPTSGAVITGHDITDEDEIRKNLGMCPQHNVLFDRLTVEHIMFWFARLKGSLSKH 120
QY 1008 VAEEMQALNDVGLPSSKLSKTSQLSGGMORLKSVALAFAGSGKVITLDEPTAGVPDYS 1067
    ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
Db 121 IRKETDKMIEDEL-SNKRHSILVOTLSCGMKRLKSVALAFAGSGRAIILDEPTAGVPDYS 179
QY 1068 RRGITELLKTRKQGTITLSTHMDADVLGDRIALISHGKLCVGSGLPKNOIGTGY 1127
    ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
Db 180 RRAIMDLILKTRKPGRTITLSTHMDADVLGDRIALISHGKLCVGSGLPKNOIGTGY 239
QY 1128 LTLVKDVESSLSCRNSSITVYLKKEEDSVSOSSDAGLGSDEHSDTLTIDVAISNLI 1187
    ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
Db 240 LTLVKQAPAPGISOEPGLASSPSCGRLSSCSQEPD-----VSQFI 279
QY 1188 RKHVSARLVEDIGHELTLYLPEAKGCAVELEFHEIDRLSDGISVYSEITTEEI 1247
    ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
Db 280 RKHVASLILVDTSTELSYILPSEAVKKGAFERLFOLEHSLDALHLSFGLMDTTELEV 339
QY 1248 FLKVAEE-----SGVDA-ETSDGTLPARRRRAFGDKQSLREPTTE----- 1287
    ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
Db 340 FLKVAEEDOSLENSBADVKESRKDVLPGAEGITAVGGAGNLARCSELSAQSASLQASASS 399
QY 1288 -----DDAADPNDSIDIPRESRETDLSGMDGSKYQVKGKML 1324
    ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
Db 400 VGSARGEGTGYSDQGYRPLFNLQDPD--NVSLQEAEMELAQY-QGOSRKLDEGMWL 456
QY 1325 TQOQVALLMKRLIARSKGFAQIYLPVAVFCIALVSLVPPRGKIPSELOPMY 1384
    ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
Db 457 KMROHGLVLRPHOARNSKALCSQIILPAFVCAVATVAVLPEIGDIPPLVSPSQY 516
QY 1385 NEQYT-----FVSNDAPE-----DTGTELLNALTKDPEFGTRCM----- 1419
    ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
Db 517 H-NTQPGRNFIPIYANERQRYRLRLSPDASPOQDVLSTFRLPSVGATCVLKSANGSLG 575
QY 1420 -----EGNFI-----PD----- 1426
    ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
Db 576 PMLNLSGESRLLAARFDSKLESTFQGLPLSNFVPPRPSAPSDSPVXPDEDSLQAWN 635
QY 1427 -TPOAGEEEMTTAP-VPOITMDFONGMNTQONPSACOCSSDKTKMLPVPFAGG 1483
    ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
Db 636 MSLEPTAPGPEITWAPSILPLVHEPR-----CTCSAQGTGFS---CPSSVGG 680
QY 1484 LPPQOKONTADILQDLTGRNISDYLVKYVOLIAKSLKNIWNEFRYGGFSLGVSNQ 1543
    ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
Db 681 -HPQMRVVTGDIILDTIGHNVSEYLLFTSDRF-----RLHRYGATITFG--NVQ 726
QY 1544 ALPQSEVNDALIKMKKHLAKDSSADRLNSIGRMTGIDTRNNVKVWFNNKGWHAIS 1603
    ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
Db 727 KSIPLAS-----FGARVPMVRAITAVRAQVILYNNKGYSMP 763
QY 1604 SFLVNNAILRANLQGE-NPSHYGITAFAHPLNLTQOOLSEVALTTSVDVLVSQVLI 1662
    ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
Db 764 TYLSLNNAILRANLPSKSKGPAAXITVYNNHNNKISASLS-LDYLLOSTDVIAFIT 822
QY 1663 FAFMFVPAFVFLIOERVSAAKHLQFTSGVKRYVLSNFWDMCNVYPATLVIIIFT 1722
    ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
Db 823 VAMFVPAFVFLVLAERSTAKHLQFVSGCNPIYVLAIVYVMDMLYLVAPCCVILIF 882
QY 1723 CFQOKSVVSTNLPVLLLLLVGWSITPLMYRASFVEKIPSTAYVVLVSVNFIQNGS 1782
    ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
Db 883 VFDLPATTSFNPFAVLSLFLVGSITPLMYRASFVEFVPSVYVFLVILNFIQTAT 942
QY 1783 VAFVLELFT-DNKLNININDILKSVLFPFHCIGRLIDVNRNOAMADALEFGE-NRF 1840
    ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
Db 943 VAFVLELFTDHEHDKLVVNSYLSKCLIFPNYNLNGHLMMAVNEYINEYVAAIGQFDKM 1002
QY 1841 VSPLSMDLVGRNLFAMAVEGVFLITVLVLIQYRFFIIRPPVNAKLSPLENDDEDRERO 1900

```

```

Db 1003 KSPFENDIYVRGLVAMTVBEEFVGLTIMQYNFLQRPRLRVSRTKPYVD-DYDVAASERQ 1061
    ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 1901 RIIDGGQNDIELEIKELTKIRRK---RRPAYDRICVGI-PRGECGILGVNGAKSSTF 1956
    ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
Db 1062 RVLRGDADNDMVKIENLTKYKSRKIGRILADVRLCLGVCPGEGCGILGVNGAKTSTF 1121
QY 1957 KMLTGPTVYRGAPFNKNSIISNIEHVONMGYCPQPAITELLTGREHVEFFALLRGV 2016
    ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
Db 1122 KMLTGESTITGGAFVNGSHVLKDLOVOOSLQYCPQFVPPVDELAREHDLITRLKCI 1181
QY 2017 PEREKGVEEMALIKGLVYKGEKAGANSGGKRRKLTAMALIGGPVYVLEDEPTTGM 2076
    ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
Db 1182 PMDEQAVVMALEKLELTKRYADKPAGTYSGGKRLSTALAILGYPAFIFIDEPTTGM 1241
QY 2077 PKARFLMNCALSVYKESVYVLTSSMECEALCTRAIMVNGRCLGSVOHLKNRFG 2136
    ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
Db 1242 PKARFLMNLIDLTKTGSVLTSSMECEALCTRAIMVNGRCLGSVOHLKNRFG 1301
QY 2137 DGYTIVRIAGSNPDLKPYQDFGLAFPGSVLKEKRRNLQYQPLSSLSLARIFSILSQ 2196
    ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
Db 1302 DGYMITVR-TKSSQNVKDVVFFNRFPAHAGKTPRYVQYQKLSHISLAQVSKMEO 1360
QY 2197 SKRLLIEDYSQTTLLDOVYVNAKQSD 2227
    ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
Db 1361 VVGVLGIEDYSQTTLLDNVFNFAKKQSDN 1391
    ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :

RESULT 4
US-08-762-500-27
? Sequence 27, Application US/08762500
? Patent No. 6030806
? GENERAL INFORMATION:
? APPLICANT: Landes, Gregory M.
? APPLICANT: Burn, Timothy C.
? APPLICANT: Connors, Timothy D.
? APPLICANT: Dackowski, William R.
? APPLICANT: Van Raay, Terence J.
? APPLICANT: Klinger, Katherine W.
? TITLE OF INVENTION: NOVEL HUMAN CHROMOSOME 16 GENES,
? NUMBER OF SEQUENCES: 83
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: GENZYME CORPORATION
? STREET: One Mountain Road
? CITY: Framingham
? STATE: Massachusetts
? COUNTRY: United States of America
? ZIP: 01701
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: PatentIn Release #1.0, Version #1.30
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/762,500
? FILING DATE: 09-DEC-1996
? CLASSIFICATION: 435
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 08/665,259
? FILING DATE: 17-JUN-1996
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: PCT/US96/10469
? FILING DATE: 17-JUN-1996
? ATTORNEY/AGENT INFORMATION:
? NAME: Dugan, Deborah A.
? REGISTRATION NUMBER: 37,315
? REFERENCE/DOCKET NUMBER: 1G5-9.3
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (508) 872-8400
? TELEFAX: (508) 872-5415
? INFORMATION FOR SEQ ID NO: 27:
? SEQUENCE CHARACTERISTICS:

```


TELECOMMUNICATION INFORMATION:
 TELEPHONE: (508) 872-8400
 TELEFAX: (508) 872-5415
 INFORMATION FOR SEQ ID NO: 25:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1684 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-665-259-25

Query Match 22.4% Score 2642.5; Db 3; Length 1684;
 Best Local Similarity 35.7% Pred. No. 1.2e-241;
 Matches 639; Conservative 300; Mismatches 523; Indels 327; Gaps 48;

534 AGVETGTTGSGIE-LPHHVKYKIMRIDIDNERTN-----KIKDGYW----- 574
 111 AAVVEFEPHNSKEPLPLAVKHLRF---SYTRRYMWTQSGFFLKTBEGMHTTSLFPL 167
 575 --DPRPADPED---MRYWGFAYLODVVEQALIRVLGTE-----KKTGYWQOMP 624
 168 FPNPBPRLTSPDGEPEYIRGFLAVOHAVRAIMEHADAATROLFQRLTITIKRPY 227
 625 PCYVDIDFLRWKSRMPLMTLAWISVAVITIKGIYVEKRAKLEKTRIMGLDMSILMFS 684
 228 PFIADPLVIAIQYQPLPLLLSTFYTALITARAAYOKEKRLLEYMMGLSMLWSA 287
 665 WEISLILLYSAGLLVILKLG-----NLPSDPSVVEFVFAVVTLLQCLISTL 739
 288 WFLLEFLLLIAASFMTLIFCVKVPNAVAVLSRSDPSLVFLFLCEALSTFSFVSTF 347
 740 FSRANLAAACGGIIFLYLPLVLCVAMQDYVGFTIKFASLSLSPVAFGECEFALE 799
 348 FSKAMMAAFSGFLYFYTYIPFEVAPRYNMWITLSQKLSCLLSNVAAKMAQILGFEA 407
 800 OGIGVQWNLFESEPYE-EDGFNLTTISMLDTEFLYGVMTVIEAVFGQYGLPRPYE 858
 408 KGMGIOWMDLL-SPYVNVDDDFCGVGLMLLIDSVLYGLVTWMEAVPQGFVQPMWF 466
 859 PCTKTYMFE-----ESDEKSHPGSNOKRMSLCEMEEPNHLKGSYQWLVYVDGM 912
 467 FIMPEYWGCKRAVAAKKEEDSDP---EKALRNEYFEPEPDVLAAGIKIKLSVVFVGN 523
 913 K--VAVDGLALNFEQQLTSLFNGHAGKTTTMSLITGLFPPTSGTAYIILKDIRSEMST 970
 524 KRAVAVRLNINLVEGQITVLIGHNAGKTTTISMLTGLFPPTSGRAYISGYEISQWYQ 583
 971 IRONIGVQPHNVFDMLTVEEHIMFYARLKLSEKHAKAMEDALDVLPSKILSKT 1030
 584 IRKSLGICPDHDLFEDNLTVAEHLIFYAOLKGLSRKOCPEEVKOMLHIGL-EDKMSRS 642
 1031 SOLSGMOKRLSVALAFYSGKVVILDEPTAGVDPYSRGITWELLKYRQGTITLSTH 1090
 643 RFLSGMRRKLSIGALIASGKVLILDEPTSGMAISRRAIWDLLQOKSRITVTLTHF 702
 1091 MDEADVLCGLRIATISHGKLCVSSSLFKNLQNGYLLVTLAKKDVESLSSCRNRSSTVS 1150
 703 MDEADVLCGLRIATISHGKLCVSSSLFKNLQNGYLLVTLAKKDVESLSSCRNRSSTVS 1150
 1151 YLKKEDEVSSQSSDAGLSGSHESDTLITDVSALINLRKHVSEKRLVEDIGHELITVLPY 1210
 747 HCNEDP-----ISOLVHHVHPNATLSSAGAEISLFLPR 780
 1211 EAKGFAVELHEIDRLSDGISYGISETTEEFLLKAE--ESGVAETS DGLPA 1268
 781 ESTHR--FEGJFAFLKLEKKOKELIASFGASITTEVEFLVNGKILVDSMDIOAQ--LPA 836
 1269 ---RNRRAFG---DKOSCLRPTEDDAADPN---SDIDPESREITLLSGMDGKGSYOV 1319
 837 LQYHERASDVAWVSNLC-----GAMPSPDGIAGILEERTAVYKNTGL----- 881
 1320 KGMKLTQOOFAALLMKRLLIARRSRKGFPAQIVLPAVFCALVFSLLVPPFGKYPSEL 1379

882 ---ALHQOQFMAFLKKAAYSWMKMAVAQVLPITCVTLALL----- 922
 1380 QPMWYNQYTFVSNDAPEDEGTIELLNAITKDPGFSTRCEMGNPIPDPTCQAGEEWTTA 1439
 923 -----ATNYSSELPDPM--RLTLG-----EYGR 946
 1440 PVQITMDLFEONGNMTMONSPACQCSSDKIKMLPVCPGAGLPPOROKONTADILQD 1499
 947 VVPFSVPGTSQLOQLSEHLKDALQAE-----QEPREVLGD 983
 1500 LTGRNISDYLKTYVOI IAKSLKNKIWNDFRY---GGFSGLVSTQALPPSQEVNDAI 1555
 984 L-----EFLI-----FRASYBGSGFN----- 1000
 1556 KQMKHKLIAKQSSADRLNSIGRFMTGLDTRNNYVWVWNGWHAISSFLVINNALIR 1615
 1001 -----ERCL--VAASFVDYGERVYVNALEFNQAYSPALAVDNLFLK 1043
 1616 ANIQKGNDSHGITA FNHP-----LNLTKQQLSEVALMTTSVDVLSICVIFAMSFPVA 1670
 1044 --LLCG---PHASIVSNFPQPSALQAKQDFNE---GRKGFIALNL--LFAMAFIAS 1093
 1671 SFVFLIOERVSARKHLOISGKPIYKLSNFVMDCMYVVPATLVITIFCFOOKSYV 1730
 1094 TETILAVSERAVQAKHVQSVGVHVASEWLSALUMLISFLIPSLILLVFEKAFDVRFT 1153
 1731 SSTNLPLVALLLLYGSITPLMPASFEVKIPSTAYVLTSVNLFIGINSATFVEL 1790
 1154 RDGMADFTLLLLLYGMAIIPLMYLMNFFLGAAATVATRTLTIFNLISG---AFLWYT 1209
 1791 ---FTDNKLNINIDILKSFELFPHCLRGILDMYKN-----QAMADALERG- 1836
 1210 IMRIPAVKVEKSTLDHFLVLPNHCIGMAVSEYENETRYCTSSVAAHYCKYNI 1269
 1837 ---ENREVSPLSMDL--VGRNLFAMAVEGVFELITVLIOYRFFTRPPVNAKL----- 1885
 1270 QYQENY-----AMSARGVRFVASMASGCAVLILFLETNLLORLICALRRRTL 1325
 1886 ---SPLNDEEDYRERORILLDGGGQNDI---LEIKELTKYIRRRKP--AVDRICV 1934
 1326 TELYTRMPVLPEDQVADRTLRILAPSPSLHTPLTIKELSKY--EQRVPLADRLSL 1384
 1935 GIPRGEFCGLGVNAGKSTFKMLTGDTTVRGCAFUNKNSILNIEHVONMGCQOF 1994
 1385 AVQKGCPLDGFENAGKTTTKMLTGESLTSNGAFVGGHNISSDVGVKVRORIGYCOF 1444
 1995 DATFELLTGREHVEFALLRGVPEREKGVEGMAIRKGLVYKGYAGNTSGNKRKLS 2054
 1445 DALLDMGTREMLVMYARLGRIGPERHIGACVENTLRLGLEPHANKLVRTYSGNKRKLS 1504
 2055 TAMALLIGSPVVEFDEPTGMDPKARRFLMNCALSVKSGRSVNLTSMSMECEALCTRM 2114
 1505 TGLALGEPVAVFLDEPSTGMDPVARRLMDTVARAREGKATITTSMSMECEALCTRL 1564
 2115 AIVNNGRFGCLGSVOHLKRFSGDYTIYVRI--AGSNPDLPVQDFGLAPPGSLAKKH 2172
 1565 AIVNNGRFGCLGSVOHLKRFSGDYTIYVRI--AGSNPDLPVQDFGLAPPGSLAKKH 2172
 2173 RNNLQYLPSSSLSLARIFSLISQSKRLHEDYSQTTLDQVFNFA 2221
 1625 QGVVHYHLPRDLISWAKVFGILEKAKKEKGYVDYVSQISLEOVFLSPA 1673

RESULT 6
 US-08-762-500-25
 Sequence 25, Application US/08762500
 Patent No. 6030806
 GENERAL INFORMATION:
 APPLICANT: Landes, Gregory M.
 APPLICANT: Burn, Timothy C.
 APPLICANT: Connors, Timothy D.
 APPLICANT: Dackowski, William R.
 APPLICANT: Van Raay, Terence J.

APPLICANT: Klinger, Katherine W.
 TITLE OF INVENTION: NOVEL HUMAN CHROMOSOME 16 GENES.
 TITLE OF INVENTION: COMPOSITIONS, METHODS OF MAKING AND USING SAME
 NUMBER OF SEQUENCES: 83
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: GENZYME CORPORATION
 STREET: One Mountain Road
 CITY: Framingham
 STATE: Massachusetts
 COUNTRY: United States of America
 ZIP: 01701
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/762,500
 FILING DATE: 09-DEC-1996
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/665,259
 FILING DATE: 17-JUN-1996
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: PCT/US96/10469
 FILING DATE: 17-JUN-1996
 ATTORNEY/AGENT INFORMATION:
 NAME: Dugan, Deborah A.
 REGISTRATION NUMBER: 37,315
 REFERENCE/DOCKET NUMBER: 165-9.3
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (508) 872-8400
 TELEFAX: (508) 872-5415
 INFORMATION FOR SEQ ID NO: 25:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1684 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-762-500-25

Query Match 22.4%; Score 2642.5; DB 3; Length 1684;
 Best Local Similarity 35.7%; Pred. No. 1.2e-241;
 Matches 639; Conservative 300; Mismatches 523; Indels 327; Gaps 48;

534 AGVFGITPGSE-LPHVYKIRMDIDNVERTN-----KIKDGW----- 574
 111 AAVFEPHNSKEPLPLAVKYHLRF---SYTRRNMTQTGSPFLKTEGWHHTSLPEPL 167
 575 --DPGRADPFED--MRYVWGCFALODVBOAIIIVLTGTE-----KKTGVYQOOMPY 624
 168 FPNPGRRELTPDGGEGGIRREGFLAVOAHVDBRAIMEYADATROLRPLVTTIKRPFY 227
 625 PCYVDIFLRVNSRMPLEMTLAMIYSAVVIKGIYKEARKETMRIMLGDSNLTWES 684
 228 PPIADPFLVAIOYQPLLLLSFTYALTALRAVVOEKERRLKEXMRMMSLIMHSA 287
 685 WFTSLIPILVSGLLVLIKLG-----NLTPYSDPSYFVFLSVFAVVTIIIOGCLISFL 739
 288 WELFLFLPLLAASEMTLLFCVAKPNVAVLSRSDPSVLAFPLCFAISTISFSEVSTF 347
 740 FSRANLAACGGIITFLVLAAMODYVGTLLKIFASLSPAVFCGCEYFALFEE 799
 348 FSNANNAFAFGGLFYFTIYFFVAPRYNMWMLSKLSCSLSNVAMAMAGOLLKFEFA 407
 800 QGIGVOMDLFESVPE-EDGFNLITISIMLFDTELYGVWYVTEAVPEQGYGIRPWFY 858
 408 KGMGIMRDLT-SPVAVDDDFCGQVGLMILLDSVLYGLTWYMEAVFPGQFVGPWFY 466
 859 PCRKSYWFGF-----ESDEKSHPSGNSOKRMSIECMEEPETHLKGVSQNLVKKYVROOM 912
 467 FIMPSTWCGKPRVAVGKREEDSDP---EKALRNEYFEAPEDLVAGIKIKHLSKVRVGN 523

913 K--VAVDGLALNFEGQITSLFGLNGAGKTTMSLTGLFPPTSGTAYILGKDIRSEMST 970
 524 KDRAAVADLNLNLYEGQITVLLGHNAGAKTTLSMLTGLFPPTSGAGYISGTYSIDQMW 583
 971 IRONLGVCPQHNVLFDMLTVEEHIMFYARLKLSEKHVAEMDOMALDVGLPSSKLSKT 1030
 584 IRKSLGCPQHDILFDNLTVAEHLFYAQKLSRQKPREVKOMLIIGL-EDKWNSSRS 642
 1031 SOLSGMQRKLSVALAVGSSKYVVIDEPTAGVDPYRSRKIMVLLKYNQGRITIIISHH 1090
 643 RFLSGMRKRLSIGIALIAGSKVILDEPTSGMDALSRAIWDLQROKSDRTVLTTHF 702
 1091 MDEADVLDGRITAIISHGKLCVSSSLFKNLQGTGYLTIVKRDVSSLSGSSNSSSTVS 1150
 703 MDEADVLDGRITAIIMAKELIOCCSSFLPKKYAGAGHMTLVKE-----P 746
 1151 YLKKEDEVSSSSDAGLGSDESPTLTIDVSAISNLRKHVSEARLVEDIGHELTVLPY 1210
 747 HCNPED-----ISQLVHNVHPATLESSAGALSFLPR 780
 1211 EAKEGAFVELFHEIDRLSDLGISYCISETTLEELFKVAB-ESGVAFESDGTLP 1268
 781 ESTHR--FEGFLAKLEKKOKELGASFGASITTWEEVFLVKGKLVDSMDIQAIQ--LPA 836
 1269 --RRNRRAFG--DKOSCLRPFTEDDAADPN--SDIPRESRETDLSGMDGKSGYOV 1319
 837 LQYHERRASDMAYDSNLC-----GAMPDSOIGALIEERTAVKLNGL----- 881
 1320 KGMKLTQOOQVALLMKRLLIARRSKGFPFOIYLPVAVFCIALVFSLIVPECKYPSLEL 1379
 882 --ALHCQEFWAMFLKAAASWREMKVAQVLPVLCVTALAL----- 922
 1380 QPMWYNEDYTFVSNDAPEDTGTLELLNALTKPDGFCRMEGNIPDTPQAGEEEMTTA 1439
 923 -----AIWYSELEFDDPML--RLTLG-----EYGR 946
 1440 PVPTQIMDLFQNGMWTQNSPACOCSSDKIKKMLPYCPGAGGLPPQKONTADILQ 1499
 947 VPFPSVPGTSGQGLDLSHLKDALQAG-----QEPREVIGD 983
 1500 LTGRNISDYLVKTYVOIIANSLSKKNIVNEFRY---GGSFLGVSNTQALPPSGOVNDAI 1555
 984 L-----EEFLI-----FRASVEGGGFN----- 1000
 1556 KQMKHLKLAODSSADRPRLNSIGRFMTGIDTRNNVKKWPNFNKGNHATISFLVYINNAILR 1615
 1001 -----ERCL--VAASFPDVGERTVVALFNNOAISHPATLAAVVDNLLEK 1043
 1616 ANLOGENSHYGITAFNHP-----LNLTKQOLSEVALMTTSVDVLVSICIFRAMSFVA 1670
 1044 --LLCG---PHASIVSNFQPRASALQAAKQFNE---GRKGFIALNL--LFAMATLAS 1093
 1671 SFVVFLEIDERSKAKHLOFISGVKPVYIWMLSNFVDMCNVVPATLVITIFICPOOKSYV 1730
 1094 TFSILAVSERAVOAKHVOQVGVHVASFWLSALMDLISFLIPSLILLVYKFAEDVAF 1153
 1731 SSTNPLVALLLLLYGMSITPLMYPASVFVKIPSTAVVLTYSVLFGISGVATFVLEL 1790
 1154 RDGHMADTLLLLLLXGNAIIPMLTLMNFFLGAATATRTLTIFILISGT---ATFLMWT 1209
 1791 --FTDNKLNINDILKSVFLIPFPCLGRGLIDMVKN-----QAMADALERFG- 1836
 1210 IMRIPAVVLEELSKTLIDHVLFLPNHICGMVSSFEVEYETRYVCTSSSEVAABHCKKYNI 1269
 1837 ---ENKREYSPLSWDL--VGRMLFAMAVEGVVFFLITVLIQYRFRIRRPVNAKL----- 1885
 1270 QYQENFY----AMSAFGVGRFVASMAGCAVYLLFLFLEIINLQRLGILCALARRRTL 1325
 1886 -----SPUNDEDEVREPORITLDGGGONDI---LEIKELTKYRRKKRP--AVDRICV 1934
 1326 TELYTRMVLPEPDQVADERIRIILAPSPDSLHTPLIKELSKYV--QORVPDLAVDLSL 1384
 1935 GIPGECFGLLVNGAGKSSTFKMLTGDITVTRGDAPLNNKNSILSNIEVHONKGYCPQF 1994

Db	1385	AVQGEQCGLLGFENFQACKITTFKMLTGEESLTJGDAFAFGVGHRISSDVGKVRQIRIGYQPF	1444
QY	1995	DAIFELLITGREHAFEPFLITGVPEKEVGKGEAMIAIKLILVYXGKGVAGNSGCKRRLS	2054
Db	1445	DALDDHMGKRLMVMYARKLGITPEPHIGACVENTITGLLEPIANKLVATYSGCKRRLS	1504
QY	2055	TAMALIGRPVFLDEPTTGMDEPKARFLMNCALSYVKEGRSVLTLSHMECECNALCYRM	2114
Db	1505	TGIALIEPAVIFLDEPSTMDPVARLLMDTVARARESGKAITTSHSMECECNALCYRL	1564
QY	2115	AIMWNGFRCLISVOHLKKNFGEGYITLVRI - AGSNPLKTVQVDFPGALPAGSYLKKH	2172
Db	1565	AIMVQGEQKICGSPQHLKSKFSGYSILRAKVQSGQQLAELEFKAFVDTLPFGSVLEDEH	1624
QY	2173	RNMLOLYQLPSSLSLAIFESILSQQSKRLHIEDIYSVSQTTLDQVFNRA	2221
Db	1625	QGMVNHILPGKIDSWAVVGVITLIEAKKEKGVVDYSVQSLIEQVFLSRA	1673


```

QY 1556 KKKKKHKLAKSDASDRFLNSGRFMTGLDPNNKVKWPNKKGMHATSSPLTNINNALR 1615
DB 1021 -----ERCL--VAASFROGERTYVNALENNQAHSPATLAAVDNLLFK 1063
QY 1616 ANLOGENSHYITAFNHP-----LNLTKOQLEVALMTTSVDVLSICVIFAMSEVPA 1670
DB 1064 --LLCG---PHASIVVSNPPQSRSLQAQKQFNE--GRGFDIALNL--LFMAAFILAS 1113
QY 1671 SFVEFLIDERSKAAHLOFISGVKVIYIWSNFVDMCNVYVATLITIFICQOKSYV 1730
DB 1114 TFSIIAVERAAQAHVQFVSGVHASFMSALMDLISFLPSLLLVFKAPDVRAFT 1173
QY 1731 SSTNPLVALLLLYGMSTPLMPASPFVKIPSTAYVVLTSVNLFGINSVAFVEL 1790
DB 1174 RCGHAAOTLLLLLYGMAITPLMTLMNFFLGAATAYTRLTIFNLISGI---ATFLVWT 1229
QY 1791 ---FTDNKLNINDILSKVLELFPHCIGRLIDVKN-----QAMADALEREFG- 1836
DB 1230 IMRIPAVKLEELSKTLDHVELVLPNHCIGMAVSSFEYENETFRYCTSSVAAHCKKYN 1289
QY 1837 ---ENRFVSPILSMDL--VGRNLFAMAVEGVVFLITVLQYFFLRPRPVAKL----- 1885
DB 1290 QYQENFY---AMSAPGVGRFVAMASACAYLLFLFIETNLQRLGILCALRRRRTL 1345
QY 1886 -----SPINDEDEDVRRRORILDGGGNDI---LEIKELTKIYRRKKRP--AVDRICV 1934
DB 1346 TELYTRMPLPDPDQVADERTILAPSPSLHPLILIKELSKV--EORVPLLAVDRLSL 1404
QY 1935 GIAPGECFGLLVNGAGSKSTFKMLGDTTTRGDAFLNKNLSILNIEHVQNMVCPQF 1994
DB 1405 AVQKGCFCGLLVNGAGSKSTFKMLGDTTTRGDAFLNKNLSILNIEHVQNMVCPQF 1464
QY 1995 DATETLLTGRENHVEFPALLRGVPEKEVGVGEMAIRKILGVYKGRKAGNYSCKKRLS 2054
DB 1465 DALDMDTREMILVMYARLGIPIERHIGACVENTLRGLLEPHANKLVRTYSGKRRKLS 1524
QY 2055 TMAALIGBPVFLDEPTGMPKARFLMNCALSVYKGRSVLTSHMECEALCTTRM 2114
DB 1525 TGAALIGBPVFLDEPTGMPKARFLMNCALSVYKGRSVLTSHMECEALCTTRM 1584
QY 2115 AIMWNGRPGCLSGVHLLKNRFGDGYTIVRT--AGSNPLKPVQDFGLAPGSVLKEKH 2172
DB 1585 AIMWNGRPGCLSGVHLLKNRFGDGYTIVRT--AGSNPLKPVQDFGLAPGSVLKEKH 1644
QY 2173 RNMLQYQLPSSLSLARIFSIISSQKKRLHIDYSQOTTLDOVVNFA 2221
DB 1645 QGNVHYHLGRDLSSWAKVFGILEKAKETGVDDYSVQSLSEOVLFSA 1693

RESULT 8
US-08-583-276-19
: Sequence 19, Application US/08583276
: Patent No. 5837536
: GENERAL INFORMATION:
: APPLICANT: McDonagh, Kevin T.
: APPLICANT: Nienhuis, Arthur
: APPLICANT: Tolstoshev, Paul
: TITLE OF INVENTION: IMPROVED EXPRESSION OF HUMAN
: TITLE OF INVENTION: MULTIDRUG RESISTANCE GENES AND IMPROVED
: NUMBER OF SEQUENCES: 19
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Cecchi & Stewart
: STREET: 6 Becker Farm Road
: CITY: Roseland
: STATE: New Jersey
: COUNTRY: USA
: ZIP: 07068
: COMPUTER READABLE FORM:
: MEDIUM TYPE: 3.5 inch diskette
: COMPUTER: IBM PS/2

```

```

: OPERATING SYSTEM: PC-DOS
: SOFTWARE: DNA V2
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/583,276
: FILING DATE: 05-JAN-1996
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/332,444
: FILING DATE: 31-OCT-1994
: APPLICATION NUMBER: 07/887,712
: FILING DATE: 22-JAN-1992
: INFORMATION FOR SEQ ID NO: 19:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1280 amino acids
: TYPE: amino acid
: STRANDEDNESS:
: TOPOLOGY: Linear
: MOLECULE TYPE:
: DESCRIPTION: protein
: US-08-583-276-19

```

```

Query Match 2.98; Score 340.5; DB 2; Length 1280;
Best Local Similarity 18.78; Pred. No. 4.1e-22;
Matches 272; Conservative 201; Mismatches 470; Indels 509; Gaps 56;

```

```

QY 776 KIFASLSPVAFGCGEYFALFEEOGIGVOMDLNFBSPVEDEDFNLTTSSIMLFTFTL- 834
DB 256 EVLAIRIVAFGQKKELELYNK-----NIEEA--KRIGKKAITANISIGAFL 305
QY 835 ---YGVWTW-IEAVPEGOYIPR---PWYFCTKSYMFGEES----- 870
DB 306 IYASVALAFVGTTLVLSGEYSIGQVLTFFSVLIGAFSGQAQSPLEAFANRGAAYEI 365
QY 871 ---DEKSHPSNCKRNSICEEPEPHLKGVSIQNLVAVYRDGMKVAV--DGLALNFE 925
DB 366 FKIIDNPKSIDYSK-----SGHKPNKGNLEFRVNHSTYSRKEVKILKGLNVLQVS 419
QY 926 GQITSFLHNGAGCTTMSILTGLEPPTSGTAYILGDIRS--EMSTIRQMLGVCPOHNL 984
DB 420 GQYVALVNSCGSGSTVOLMORLYDPTBEGVNSVDGDDITINRFLRELIGVSGQPV 479
QY 985 FDMLTVEHIFVYRLKGLSEKHVKAEMQMALD--VGLP---SSKLSKTSQLSGGMOR 1039
DB 480 F-ATTIENI--RYGRONVYMDIEKAKVAMAYDFIMKLPKEDTLVGERGAQLSGQOK 537
QY 1040 KLSVALAPVSGSKVYIIDEPYAGVDPYSRGIWELLLKYROGRTIISTHMD---EADV 1096
DB 538 KIALARALVNRPKTLLDDEATSLADTESAVVOVALDKARKRTTYIAHRLSTVRNADY 597
QY 1097 LG--DRIAIISHGKLCVGSLSFLKNQLGTYLTVYKKDV---ESSLSGRNSSSTVSY 1151
DB 598 IAGFDGIVIEKG-----NHDELKMKK--GIYFKLVMTWQAGNEVELEMAADESKS--- 646
QY 1152 LKEDSVSOSSDAGLGSDDHESDITLIDVSAISNLIRKHSEARLVEDIGHETLYVLYPE 1211
DB 647 --EIDALEMSSNDR-----SSLIRK----- 665
QY 1212 AAKBGAFFELHEIDRLSDLGISSTGSETTLEIFLKAABESGVDAERTSDGLPARRN 1271
DB 666 -----RST 668
QY 1272 RRAFGDKQSLRPFTEDDAADPNDSIDPESRETLLSGMDKSGSYOVKGKLTQOQFVA 1331
DB 669 RRSVAGSQADQRKLSTKCAL--DESIPP-----VSFWIRIKLMLTE 707
QY 1332 LMKRLLIARRSKRGFPQOYVLPVAVFCIALVFSLIYPPFGKYSLELDQVMNTEQYTFV 1391
DB 708 --WPYFVV-----GVFCALINGGIQAPAFATIFSKII----- 736
QY 1392 SNDAPEDVTGLELLNALIKDKDGFGRMGEGNPIPDTPQAGEEMTTAPVPOTIMDLFON 1451
DB 737 -----GVFTR----- 741

```

QY 1452 GNMWQNPSPACOCSSDKIKKMLPVCPRGAGLPPRQKONTADILLODTGNISDIYLVK 1511
 Db 742 -----IDDETRKRONSNLFSL-----FLAL 762
 QY 1512 TVVQIIAKSLKKNKIMWNEFRYGSISGVSNTQALPPSQOVNDAIQMKHKLAKDSS-A 1570
 Db 763 GIISFT-----FLOGFTFGKA-----GELLTKRLKRYMFRSLKRDVSWF 804
 QY 1571 DRELNSLGRPMGTDRNNVVKWFNNKGMHAISSEFLVNNALRANLOKGENPSHYGT 1630
 Db 805 DDPKNTGALTTRL-ANDAQV---KG--AIGSLAVITONI--ANLGTG-----II 849
 QY 1631 AFNHPLNLKQOLSEVALMTTSVDLVSLCVIFAMSEVPASVVLIDERSKAKHLOFI 1690
 Db 850 SFTYGWLT-----LILLAIVPIIAGVVEK-----MISQALDKKKELE-- 891
 QY 1691 SGKPVYIWLNSFVDMCMYVVPATLVIIIFICPOK---SVYSTNIPVALL--LLLY 1745
 Db 892 GAGKIATEAIEFN-----RTVVSILTOQOKFEHMAQSLQVPRNSLKAHIF 938
 QY 1746 G--WSTTPMTYASFEVKLPSTAYVVLTSVNLFIGINSVATFVLEFTDNKLNINDIL 1803
 Db 939 GTFSETOAMMYFSAQCFRFGAYLVA-----HKLMSEFEDVL 975
 QY 1804 KSEVFLIFPHFCRGLIDKVKQAMADALEPGENRFV-----SPLSMDIVGRNLFAMAY 1858
 Db 976 ----LVFSAVVEGAMAVGVSSFADYAKAKISAHIIIMIEKTPILDSYSTEGLMPVTL 1031
 QY 1859 EGVVEFLITVLIIQREFTRP-PVNAKLSPLNDEDEDERERORILLDGGONDIIEIKEL 1917
 Db 1032 EGNVTEGEVY---FMYPRPDIPVLOGLS-----LEVKK- 1062
 QY 1918 TTYRRKKRPAYDRICVGPPECEGILGVNAGKSTFEKMLTGTITVTRQDAPLANKSI 1977
 Db 1063 -----GQTLALVSSGSGKSTVOLLERFVDPILACKVLLDCKEI 1101
 QY 1978 LS-NIHEVQNMGYCPQ---FD-----ATTELLTGREHEFFALLRGV 2016
 Db 1102 KRLNOMLRAHLGIVSQERILFDCSIAENIAYGDSRVVSOEIVRAKKNIAHFIESTL 1161
 QY 2017 PEKEVGKVEAMIRKLGIVKYEKTAAGNYSGGNKRKLSTAMALIGCPVPLDEPTGMD 2076
 Db 1162 PKRYSTKVGDKTQ-----LSGQOKORIAIARALVQPHILLDEATSMALD 1207
 QY 2077 PARRELMNCALSVYKESRNVLTSHSMECECALCTRMAIWNNGFRCLGVOHLKKNFG 2136
 Db 1208 TESEKVVGE-ALDKAREGRTCIIVAHRLSTION-ADLIVFQNGKVEKHTHQQLLAOKG 1265
 QY 2137 DGYTIVRIAGS 2148
 Db 1266 IYFSMVSVQAGT 1277

RESULT 9

5206352-4

Patent No. 5206352

Applicant: Roninson, Igor B.; Pastan Ira H.; Gottesman, Michael M.

SEQUENCES ASSOCIATED WITH MULTIDRUG RESISTANCE IN HUMAN CELLS

TITLE OF INVENTION: COMPOSITIONS FOR CLONES CONTAINING DNA

NUMBER OF SEQUENCES: 4

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/622,836

FILING DATE: 24-SEP-1990

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 892,575

FILING DATE: 01-AUG-1986

APPLICATION NUMBER: 845,610

FILING DATE: 28-MAR-1986

SEQ ID NO: 4

LENGTH: 1280

5206352-4

Query Match 2.9%; Score 339.5; DB 6; Length 1280;
 Best Local Similarity 18.7%; Pred. No. 5, 1e-22;
 Matches 272; Conservative 201; Mismatches 470; Indels 509; Gaps 56;

QY 776 KIFASLLSPVAFGGCCGYFALFEEOGIGVOWMDLFSPEVDEDFNLTTSSIMLFTPL- 834
 Db 256 EVLAIKRTVIAFGQKKELERYNK-----NLEEA--KRIGIKAIYANISIGAAFL 305
 QY 835 ----YGYMTWY-IEAVPGQYGIPIR--PWYFPGTSYMFGEES----- 870
 Db 306 IYASVALAFWNGTTLVLSGEXSLIGQVLTFVFSVLIGAFSGQASPIEAFANRGAAYEI 365
 QY 871 ----DEKSHPGSNOKRMSSEICMEDEPTHLKGVSIONLVKVVQDKKVVAV-DGLALNFEY 925
 Db 366 FKIIDNKSIDSYSK-----SGHKPDNIKGNLEFRNHSYPSRREKVALKGLNKLKQVS 419
 QY 926 GQTSFLHGAGKATTTMSILUTGLEPTSTAYILKDIIRS-EMSTIRONLVCPOPHNL 984
 Db 420 GQYVALVGNSGCGKSTTVOLMORLYDPTBGMVSDODLTTIVRPLRETIQVSGQEV 479
 QY 985 FDMLTVEHITWYFARLKLSEKHVKAEMQALD--VGLP---SKYKSKTSQSGMOR 1039
 Db 480 F-ATTIAENI-RYGRENVITDEIEKAVKEMAVDFIMKLPKHPDTLVGRGAQLSGQGQ 537
 QY 1040 KLSVALAFVGSKVVLIDPTAGVDYPSRKGITWELLKTRQGTITLSTHND--EADY 1096
 Db 538 RIATARLAVNRPKLTLLDEATISALDTESEAVVOVALDKARKGRTTIVIAHRLSTVARNADV 597
 QY 1097 LG--DRIATISHKCLCVSGSLFLKNOLGTGYLTLYVKRDV---ESSLSGRRSSSTVSY 1151
 Db 598 IAGDDQVIEYKG-----NDELMEKER---GIYFKLYVTQTAGNEVELMAADESKS--- 646
 QY 1152 LKREDSVQSSDAGLSGDSHESDPTLTIDVASISNLRKHVSEARVEDIGHETLYVPE 1211
 Db 647 --EIDALEMSSNDSR-----SSLIRK----- 665
 QY 1212 AAKEGAVELFHEIDRLSDLGISVGISSETTLEELFKVAEESGVDAETSDGTLPARN 1271
 Db 666 -----RST 668
 QY 1272 RRAFGRQSCIRPTEDDADPNDSIDPESRETDLLSGMDKSGYQVGMKLTQOQFVA 1331
 Db 669 RRSVRSQAQDRKLTSLRKAL--DESLPP-----VSFWMIMKLNLTLE 707
 QY 1332 LMKRLLIARSKKGFPAQIVLPAVVCIALVPSLIVPPFGKTPSLDELQPMYINQYTFV 1391
 Db 708 --WPYFVY-----GVFCALINGILOPAPALITFSKLT----- 736
 QY 1392 SNDARPDGTLELLNALTKDPGFGTRCMENPIPDTPCQAGEEMTTAPVQTIMDLPON 1451
 Db 737 -----GVFTK----- 741
 QY 1452 GNMWQNPSPACOCSSDKIKKMLPVCPRGAGLPPRQKONTADILLODTGNISDIYLVK 1511
 Db 742 -----IDDETRKRONSNLFSL-----FLAL 762
 QY 1512 TVVQIIAKSLKKNKIMWNEFRYGSISGVSNTQALPPSQOVNDAIQMKHKLAKDSS-A 1570
 Db 763 GIISFT-----FLOGFTFGKA-----GELLTKRLKRYMFRSLKRDVSWF 804
 QY 1571 DRELNSLGRPMGTDRNNVVKWFNNKGMHAISSEFLVNNALRANLOKGENPSHYGT 1630
 Db 805 DDPKNTGALTTRL-ANDAQV---KG--AIGSLAVITONI--ANLGTG-----II 849
 QY 1631 AFNHPLNLKQOLSEVALMTTSVDLVSLCVIFAMSEVPASVVLIDERSKAKHLOFI 1690
 Db 850 SFTYGWLT-----LILLAIVPIIAGVVEK-----MISQALDKKKELE-- 891
 QY 1691 SGKPVYIWLNSFVDMCMYVVPATLVIIIFICPOK---SVYSTNIPVALL--LLLY 1745
 Db 892 GAGKIATEAIEFN-----RTVVSILTOQOKFEHMAQSLQVPRNSLKAHIF 938
 QY 1746 G--WSTTPMTYASFEVKLPSTAYVVLTSVNLFIGINSVATFVLEFTDNKLNINDIL 1803

```

Db 939 GTFSTQAMMFYSYAGCFRCGAYLVA-----HKIMSFEDVL 975
Qy 1804 KSVFLIFPHFCGRGLIDMKQAMADALERTGENRFV-----SPLSMDLVGRNLFAMAV 1858
Db 976 ---LVFSAVVEGAMAVGVSSSFADPYAKAKISAAHIIIEIKTPIIDISTEGIMPRTL 1031
Qy 1859 EGVFELLVLVLYQRFPIPR-PVNAKLSPLNDEDEDEVREHORILLDGGGNDIEIKEL 1917
Db 1032 EGVNVEGVV---FNVPRPDIPVLOGLS-----LEVKK- 1062
Qy 1918 TKYRRKKRPVADRICVGIPEGEFCGLLVNGAGKSSIFKMLTGTPTVTRGDAPLANKNSI 1977
Db 1063 -----GQTLALVSSCGSKSTVQLLEFYDPLAGKVLIDGKEI 1101
Qy 1978 LS-NIHEVQNMNGYCPQ---PD-----AITELLTGHEHVEFFLLAGV 2016
Db 1102 KRLNVOMLRAHLGIVSOEPIIFDCSIAENIAYGDSRVVSOEIVRAKEANITIAFISL 1161
Qy 2017 PEKEVGKVEGEMARLGLVYKGEKYAGNYSGNKRRKSLTAMALIGCPVFLDEPTTGM 2076
Db 1162 PKYSTRKVDKGTQ-----LSGQKORIALARLAVQPHILLDEATSLD 1207
Qy 2077 PARPRLMNCALSVYKEGHSVLTSHSMECEALCTRAIMVNGRPFCLGVOHLKNRF 2136
Db 1208 TESEKVVQD-ALDKAREGRTCIIVIAHRLSTION-ADLIVFQNGHVKEHGTQOLLAKG 1265
Qy 2137 DGTIVVRIAGS 2148
Db 1266 IYFSMVSVQAGT 1277

```

RESULT 10

US-08-784-649A-2
Sequence 2, Application US/08784649A

Patent No. 5830697
GENERAL INFORMATION:
APPLICANT: SIKIC, Branimir I
APPLICANT: Chen, Gang
TITLE OF INVENTION: P-GLYCOPROTEIN MUTANT RESISTANT TO
TITLE OF INVENTION: CYCLOSPORIN MODULATION
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson
STREET: 2200 Sand Hill Road
CITY: Menlo Park
STATE: CA
COUNTRY: USA
ZIP: 94025
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/784,649A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Sherwood, Pamela J
REGISTRATION NUMBER: Reg. No. 5830697 36,677
REFERENCE/DOCKET NUMBER: 06037/007001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-322-5070
TELEFAX: 415-854-0875
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1279 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-784-649A-2

Query Match 2.9%; Score 337; DB 2; Length 1279;
Best Local Similarity 18.7%; Pred. No. 8.7e-22;
Matches 272; Conservative 200; Mismatches 471; Indels 508; Gaps 56;

```

Qy 776 KIFASLSVPAFGGCEYFALFEEQIGVQVONLFESEVDEDFNLTTISMLPFTFL- 834
Db 256 EVLAIRITVIAFGQKKELERYNK-----NLEEA--KRIGIKRAIVANISIGAFL 305
Qy 835 ---YGMWTY-IEAVFPQGYIPR--PWYFRCTKSVWEEES----- 870
Db 306 IYASVALAFWGTTLVLSGEYSIGQVLTFLVSLGASVGOASPSLEAFANGAAYEIP 365
Qy 871 ---DEKSHPGSNOKRMEICMEEPETHLKLGVSTIONLVKYRQGMKAVV-DGLALNFEY 926
Db 366 KIIDNKSIDSYSK-----SGHKPRDNIKNLEFRNVHFYPSRKEVKILKGLNLVQSG 419
Qy 927 QITFLGHNGAGKTTYSILTGLFPPYSGTAYILGKDIRS-EMSTYRONLGVCPQHNVLF 985
Db 420 QTVLVGNSGCGSKSTVQVLMQRLYDPREGWVSDGQDIFTINVRFLREILIGVSOEPI 479
Qy 986 DMLTVEEHIMFYARLKLGESEKHVKAEMQALD--VGLP---SSKLKSTQSLSGGMOK 1040
Db 480 -ATTIAENI-RYGRENVYMDIEKAIVEANAYDITMKLPKFDLVGEKQALSGGQOK 537
Qy 1041 LSVLAEVGGSKVYILDEPTAGVDPYSRGIWELLTKYRGRTILSTHMD--EADVL 1097
Db 538 IAIARLAVRNPKIILLDEATISALDTESEAVYQVALDKARGRTIYIAHLSTVRYNADVI 597
Qy 1098 G--DRIAIISHGKLCYGVSSFLFKNLQGTGYTLTVKKVD---ESSLSGRSSSTVSYL 1152
Db 598 AGFDDGVYVEKG---NHDELMEK--GIYFKLVMTQAGNEVELDENADESKS---- 645
Qy 1153 KKEDSVSSSSDAGLGSDHESDTLTIYSAISNLIRKHVSARLVEDIGHETLYVLYEYA 1212
Db 646 -EIDALEMSSNDSR-----SSLIRK----- 664
Qy 1213 AKGAFVELFHEIDRLDLSIGSYGISETTLIEIFLKVAVEESGVDAETSDGTLPARNR 1272
Db 665 -----RSTR 668
Qy 1273 RAFGDKQSLRPFTEDDAADPNDSIDIPESRETDLSGMDGKGYOVKMKLQOQVAL 1332
Db 669 RSVRSQAODRKLSTKEAL--DESIPP-----VSFWRIMKLNLTE- 706
Qy 1333 LMKRLLIARRSRKGPFAQIVLPAVFVCAIVFSLIVPFCKYPSLELQPMYNNQYTFVS 1392
Db 707 -WPTFVV-----GVFCATINGGLQPAFAIIFSKIT- 735
Qy 1393 NDAPEDTGLELNLALTKDPGFGTRCMEGNIPDTPQAGEEEMTAPVQTIIDLFPQNG 1452
Db 736 -----GVFTR----- 740
Qy 1453 NWTMONPSPACQCSSDKIKKMLPVCPRGAGLPPRPORRONTADILQDLTGRNISDYLVKT 1512
Db 741 -----IDDPETKRQNSNLFSL-----FALG 762
Qy 1513 YVQIIAKSLKKNKIWNNEFRYGGFSLGVSNQALPPSQGEVNDAIKQMKHLKLAKDS-AD 1571
Db 763 ITSFIT-----FFLOGFTFGKA-----GELTLKRLXYMFRSKMLRQDVSWPD 804
Qy 1572 RFLNSLGRFMTGLDTRNNKVKWFNKGMHAISFLVNNALINLANTLOKGENPSHYGITA 1631
Db 805 DPKNTTGALTTRL-ANDAAQV---KG-AIGSRLAVITQNI--ANIGTGL-----IIS 849
Qy 1632 FNHPLNLTKOOLSEVALMTTSDVNLVLSICVTFAMSEVPASVYVPLIDERSKAKHLOPIS 1691
Db 850 FLYGMQLT-----LHLLAIVPIIALAGVENK-----MLSQALKDKKKELE--G 891
Qy 1692 GVKPVIYWLNSNFWMDKMYVVPATLVLIIFICFOOK---SYVSSTNLPVALL--LLYG 1746
Db 892 AKGIATEIENF-----RTYVSLTORCKFEMHYAQSLQVRYNRSKLKAHIFG 938

```


Fri Mar 7 09:16:52 2003

us-09-595-526c-2.raii

Page 16

[illegible]

```

D0 1173 LSGGOKOPIAARALNQPILLDETSALDSEKVV0E-ALDKAREGRTCIYAHRL 1231
QY 2105 EEECALCTRAAMWNGRFRGJCSVOHILKRNFGGYIV 2142
D0 1232 STION-ADLIVLONQGVKHEGHTHOOLLAKGITYFSMV 1268

```

```
Search completed: March  7, 2003, 08:51:45
Job time : 73 secs
```

```

Db      1173  LSGGQKORIAIARAALVROPHILLDEAT$ALDTESEKVOE~ALDKARRGRTCIYIAHRL 1231
      |||:::| ||: |:: |||:::|:::| | |::|||:::|:
QY      2105  EECEALCTRMALMVNGRFRCIGSVOHKKNRFGDGYTIV 2142
      :::||:|:| | |::| |::|
Db      1232  STION~ADLIVIONCGVKEHGT HQOLLAQKGIYFSMV 1268
      :::||:|:| | |::| |::|

```


GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 7, 2003, 08:41:19 ; Search time 41 seconds
(without alignments)
2287.268 Million cell updates/sec

Title: US-09-595-526C-2

Perfect score: 11797
Sequence: 1 MACWPQRLRLMLKMLTFRRR.....VDVAVLTSFLDEKYEKESYV 2261

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues
Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

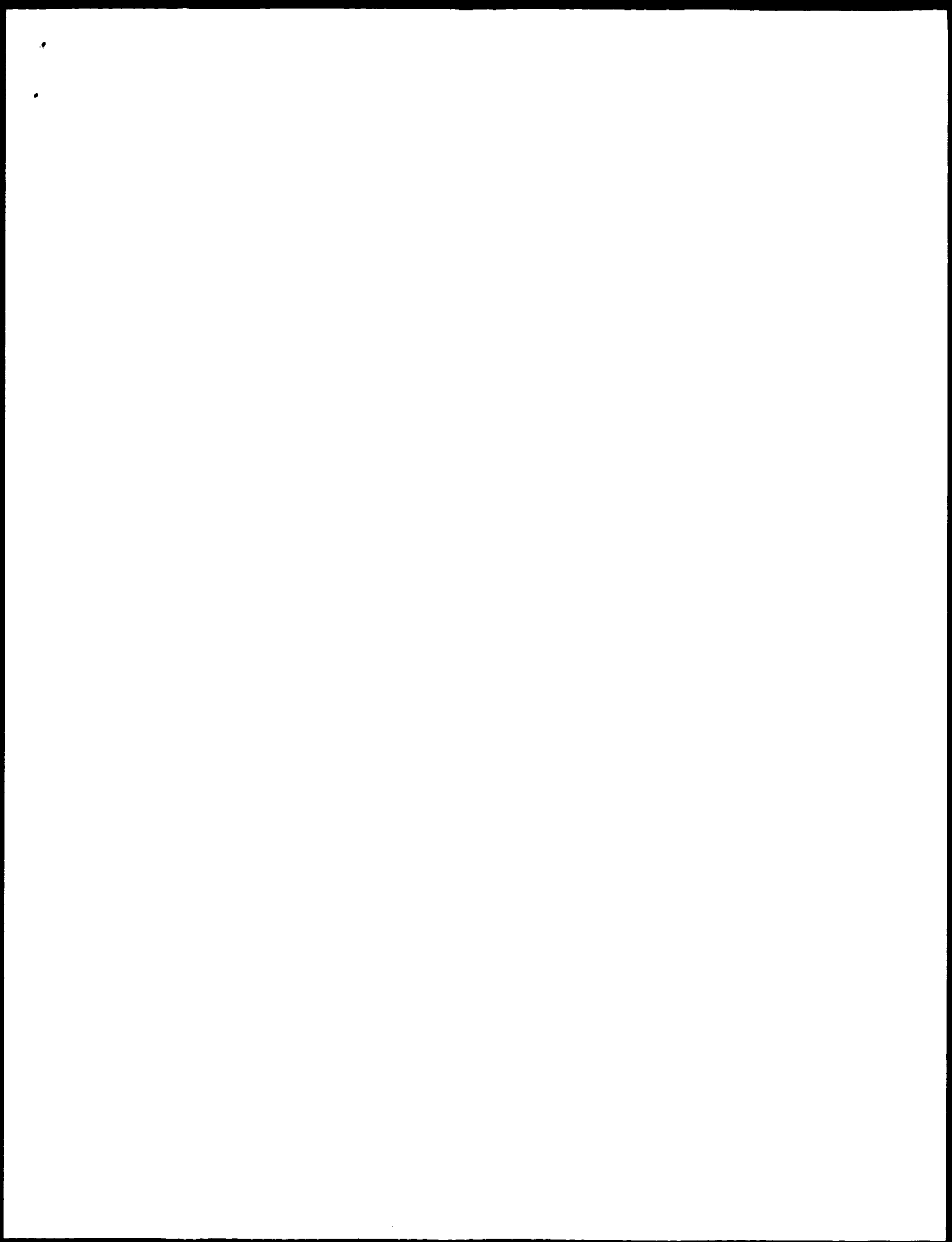
SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	11770	99.8	2261	1	ABCI_HUMAN
2	11256	95.4	2261	1	ABCI_MOUSE
3	5870.5	49.8	2273	1	ABCR_HUMAN
4	4230.5	35.9	2436	1	ABC2_HUMAN
5	4093.5	34.7	2434	1	ABC2_MOUSE
6	2642.5	22.4	1704	1	ABC3_HUMAN
7	1538.5	13.0	1704	1	CEB7_CAEL
8	411	3.5	330	1	DRRA_STRPE
9	382.5	3.2	343	1	NOD1_RHISN
10	366	3.1	304	1	NOD1_ECOLI
11	347	2.9	308	1	YADG_ECOLI
12	344.5	2.9	340	1	NOD1_RHIC
13	343.5	2.9	347	1	NOD1_RHIC
14	339.5	2.9	1280	1	MDR1_HUMAN
15	331.5	2.8	355	1	NOD1_RHIC
16	329.5	2.8	578	1	YBHF_ECOLI
17	327.5	2.8	1281	1	MDR3_CRIGR
18	327	2.8	894	1	YH1H_ECOLI
19	326	2.8	1276	1	MDR3_MOUSE
20	325.5	2.8	308	1	NOSF_PEST
21	324.5	2.8	1276	1	MDR2_MOUSE
22	322.5	2.7	306	1	MDR1_BRAJA
23	321	2.7	1276	1	MDR1_CRIGR
24	317	2.7	1362	1	PMD1_SCHPO
25	316	2.7	1276	1	MDR2_CRIGR
26	315.5	2.7	311	1	NOD1_RHILV
27	315	2.7	1278	1	MDR2_RAT
28	314.5	2.7	1277	1	MDR1_RAT
29	313	2.7	1276	1	Y415_SYNY3
30	310.5	2.6	1276	1	MDR1_MOUSE
31	304.5	2.6	1284	1	YOH5_YEAST
32	300	2.5	381	1	OPBA_BACSU
33	298.5	2.5	262	1	YA23_METUA

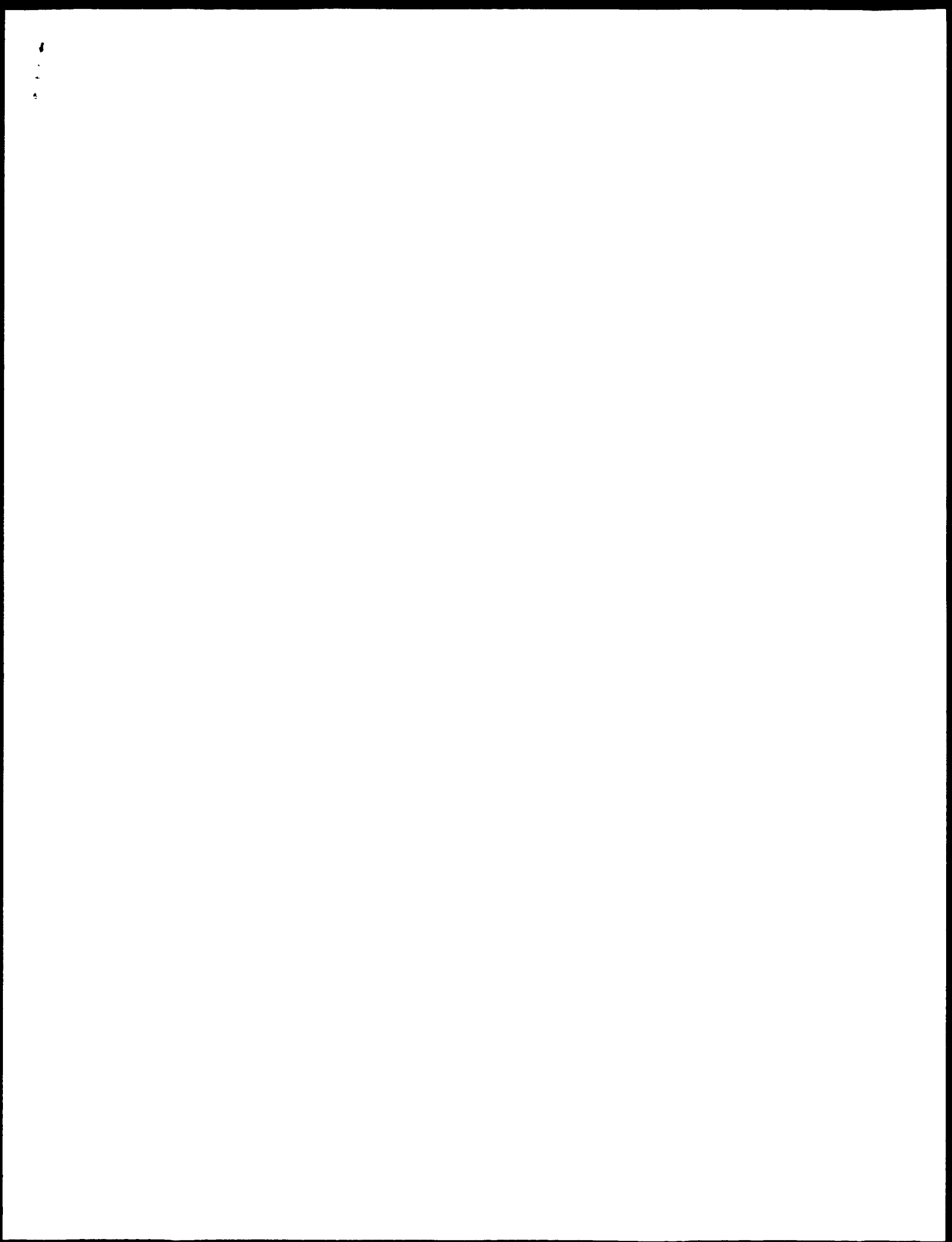
34	297	2.5	380	1	OPCA_BACSU
35	296	2.5	335	1	Y719_ANASP
36	293.5	2.5	1321	1	AB11_HUMAN
37	292.5	2.5	305	1	YHCH_BACSU
38	287.5	2.4	274	1	Y179_MYCPN
39	286.5	2.4	274	1	Y179_MYCCE
40	285.5	2.4	1279	1	MDR3_HUMAN
41	284	2.4	306	1	BCRA_BACLI
42	284	2.4	1321	1	MDR1_CAEL
43	284	2.4	1336	1	MAMI_SCHPO
44	281.5	2.4	308	1	YEHX_ECOLI
45	281.5	2.4	343	1	ABC_ECOLI

ALIGNMENTS

RESULT 1
ABCI_HUMAN STANDARD; PRT; 2261 AA.
AC 095477; Q9UN08; Q9UN07; Q9UN06; Q9UN04; Q9UN09; Q96785; Q96556;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ATP-binding cassette, sub-family A, member 1 (ATP-binding cassette transporter 1) (ATP-binding cassette 1) (ABC-1) (Cholesterol efflux regulatory protein).
GN ABC1 OR ABC1 OR CERP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
NCBI_TaxID=9606;
RX MEDLINE-20345099; PubMed-10884428;
RA Santamarina-Fojo S., Peterson K.M., Knapper C.L., Qiu Y., Freeman L.A., Cheng J.-F., Osorio J., Remaley A.T., Yang X.-P., Haudenschield C.C., Prades C., Chimali G., Blackmon E.E., Francols T.L., Duverger N., Rubin E.M., Rosier M., Deneffe P., Fredrickson D.S., Brewer H.B. Jr.;
RT "Complete genomic sequence of the human ABCA1 gene: analysis of the human and mouse ATP-binding cassette A promoter.";
RC Proc. Natl. Acad. Sci. U.S.A. 97:7987-7992(2000). -544.
RP TISSUE=Skin;
PC Schwartz K., Lavin R.M., Wade D.P.;
RA "ABCA1 gene expression and apoA-I-mediated cholesterol efflux are regulated by LXR.";
RT Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE-21251004; PubMed-11352567;
RA Qiu Y., Caveller L., Chiu S., Yang X., Rubin E., Cheng J.-F.;
RT "Human and mouse ABCA1 comparative sequencing and transgenesis studies revealing novel regulatory sequences.";
RN Genomics 73:66-76(2001).
[4]
RP SEQUENCE FROM N.A.
RX Tanaka A.R., Abe-Dohmae S., Arakawa R., Sadaami K., Kidera A., Kikoa N., Amachi T., Yokoyama S., Ueda K.;
RA "A new topological model of functional human ABCA1-signal peptide cleavage and glycosylation of a large extracellular domain.";
RT Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
[5]
RP SEQUENCE OF 21-2261 FROM N.A.
RX MEDLINE-99194549; PubMed-10092505;
RA Langmann T., Klucken J., Reil M., Liebisch G., Luciani M.F., Chamin G., Kaminski W.E., Schmitz G.;
RT "Molecular cloning of the human ATP-binding cassette transporter 1 (ABCA1): evidence for sterol-dependent regulation in macrophages.";
RN Biochem. Biophys. Res. Commun. 257:29-33(1999). [6]



RP SEQUENCE OF 21-2261 FROM N.A.
RX MEDLINE-99364413; PubMed-10431238;
RA Rust S., Rosier M., Funke H., Real J., Amoura Z., Piette J.-C.,
RA Delzenne J.-F., Brewer H.B., Duverger N., Deneffe P., Assmann G.;
RT "Tangier disease is caused by mutations in the gene encoding
RT ATP-binding cassette transporter 1.";
RL Nat. Genet. 22:352-355(1999).
RN [17]
RP VARIANTS FHA THR-1091 AND 1893-GLU-ASP-1894 DEL.
RX MEDLINE-20001430; PubMed-10533863;
RA Marcell M., Brooks-Wilson A., Clee S.M., Roomp K., Zhang L.-H., Yu L.,
RA Collins J.A., van Dam M., Molhuizen H.O.F., Loubser O.,
RA Ouellette B.F.F., Sensen C.W., Fichter K., Mott S., Denis M.,
RA Boucher B., Pimstone S., Genest J. Jr., Kastelein J.J.P., Hayden M.R.;
RT "Mutations in the ABCA1 gene in familial HDL deficiency with defective
RT cholesterol efflux.";
RL Lancet 354:1341-1346(1999).
RN [8]
RP VARIANTS TD ARG-597, AND ARG-1477, AND VARIANT FHA LEU-693 DEL.
RX MEDLINE-99364411; PubMed-10431236;
RA Brooks-Wilson A., Marcell M., Clee S.M., Zhang L.-H., Roomp K.,
RA van Dam M., Yu L., Brewer C., Collins J.A., Molhuizen H.O.F.,
RA Loubser O., Ouellette B.F.F., Fichter K., Ashbourne-Excoffon K.J.D.,
RA Sensen C.W., Scherer S., Mott S., Denis M., Martindale D.,
RA Frohlich J., Morgan K., Koop B., Pimstone S., Kastelein J.J.P.,
RA Hayden M.R.;
RT "Mutations in ABCA1 in Tangier disease and familial high-density
RT lipoprotein deficiency.";
RL Nat. Genet. 22:336-345(1999).
RN [9]
RP VARIANTS TD SER-590; SER-935 AND VAL-937, AND VARIANTS ALA-399 AND
RX MET-883
RX MEDLINE-99364412; PubMed-10431237;
RA Bodzioch M., Orso E., Klucken J., Langmann T., Boettcher A.,
RA Diederich W., Drobnik W., Barlage S., Buechler C.,
RA Porsech-Oezguermez M., Kaminski W.E., Hahmann H.W., Oette K.,
RA Rothe G., Aslanidis C., Lackner K.J., Schmitz G.;
RT "The gene encoding ATP-binding cassette transporter 1 is mutated in
RT Tangier disease.";
RL Nat. Genet. 22:347-351(1999).
RN [10]
RP VARIANTS TD ILE-929; ARG-597 AND ARG-1477, AND VARIANTS FHA LEU-693
RX DEL; THR-1091; 1893-GLU-ASP-1894 DEL AND LEU-2150.
RX MEDLINE-20540002; PubMed-1086027;
RA Clee S.M., Kastelein J.J.P., van Dam M., Marcell M., Roomp K.,
RA Zwarts K.Y., Collins J.A., Roelants R., Tamawana N., Stult C.,
RA Suda T., Ceska R., Boucher B., Rondeau C., Desouch C.,
RA Brooks-Wilson A., Molhuizen H.O.F., Frohlich J., Genest J. Jr.,
RA Hayden M.R.;
RT "Age and residual cholesterol efflux affect HDL cholesterol levels and
RT coronary artery disease in ABCA1 heterozygotes.";
RL J. Clin. Invest. 106:1263-1270(2000).
RN [11]
RP VARIANTS TD ASN-1289 AND HIS-1800.
RX MEDLINE-20171564; PubMed-10706591;
RA Brousseau M.E., Schaefer E.J., Dupuis J., Eustace B.,
RA van Berdevegh P., Goldkamp A.L., Thurston L.M., Fitzgerald M.G.,
RA Yasek-Mckenna D., O'Neill G., Eberhart G.P., Wellfendach B.,
RA Orlovas J.M., Freeman M.W., Brown R.H. Jr., Gu J.Z.;
RT "Novel mutations in the gene encoding ATP-binding cassette 1 in four
RT Tangier disease kindreds.";
RL J. Lipid Res. 41:433-441(2000).
RN [12]
RP VARIANT TD ASP-1046, VARIANT FHA CYS-230, AND VARIANTS LYS-219;
RX ILE-825; MET-883 AND LYS-1587.
RX MEDLINE-20396633; PubMed-10938021;
RA Wang J., Burnett J.R., Near S., Young K., Zimman B., Hanley A.J.G.,
RA Connelly P.W., Harris S.B., Heggele R.A.;
RT "Common and rare ABCA1 variants affecting plasma HDL cholesterol.";
RT Arterioscler. Thromb. Vasc. Biol. 20:1963-1989(2000).
RN [13]
RP VARIANT TD TRP-587, AND VARIANT LEU-2168.
RX MEDLINE-21157002; PubMed-11257260;
RA Bertolini S., Pisciotto L., Seri M., Cusano R., Cantafora A.,
RA Calabresi L., Franceschini G., Ravazzolo R., Calandra S.;
RT "A point mutation in ABCA1 gene in a patient with severe premature
RT coronary heart disease and mild clinical phenotype of Tangier
RT disease.";
RL Atherosclerosis 154:599-605(2001).
RN [14]
RP VARIANTS LYS-219; MET-883 AND ASP-1172.
RX MEDLINE-21157003; PubMed-11257261;
RA Brousseau M.E., Bodzioch M., Schaefer E.J., Goldkamp A.L., Kielar D.,
RA Probst M., Orlovas J.M., Aslanidis C., Lackner K.J.,
RA Bloomfield Rubins H., Collins D., Robins S.J., Wilson P.W.F.,
RA Schmitz G.;
RT "Common variants in the gene encoding ATP-binding cassette transporter
RT 1 in men with low HDL cholesterol levels and coronary heart disease.";
RL Atherosclerosis 154:607-611(2001).
RN [15]
RP VARIANT TD LEU-1506.
RX MEDLINE-21369429; PubMed-11476961;
RA Lapack-Bodzioch K., Bodzioch M., Kruehl M., Kielar D., Probst M.,
RA Kiec B., Andrikovics H., Boettcher A., Hubacek J., Aslanidis C.,
RA Sutrop N., Schmitz G.;
RT "Homogeneous assay based on 52 primer sets to scan for mutations of
RT the ABCA1 gene and its application in genetic analysis of a new
RT patient with familial high-density lipoprotein deficiency syndrome.";
RL Biochim. Biophys. Acta 1537:42-48(2001).
RN [16]
RP VARIANTS TD ASN-1289 AND TRP-2081, AND VARIANT LYS-219.
RX MEDLINE-21369433; PubMed-11476965;
RA Huang W., Moriyama K., Koga T., Hua H., Ageta M., Kawabata S.,
RA Nawatari K., Imanura T., Eto T., Kawamura M., Teramoto T., Sasaki J.;
RT "Novel mutations in ABCA1 gene in Japanese patients with Tangier
RT disease and familial high density lipoprotein deficiency with
RT coronary heart disease.";
RL Biochim. Biophys. Acta 1537:71-78(2001).
RN [17]
RP VARIANTS LYS-219; ALA-399; MET-771; PRO-774; ASN-776; ILE-825;
RX MET-883; ASP-1172; LYS-1587 AND CYS-1731.
RX MEDLINE-2138379; PubMed-11238261;
RA Clee S.M., Zwinderman A.H., Engert J.C., Zwarts K.Y.,
RA Molhuizen H.O.F., Roomp K., Jukema J.W., van Wijland M., van Dam M.,
RA Hudson T.J., Brooks-Wilson A., Genest J. Jr., Kastelein J.J.P.,
RA Hayden M.R.;
RT "Common genetic variation in ABCA1 is associated with altered
RT lipoprotein levels and a modified risk for coronary artery disease.";
RL Circulation 106:1198-1205(2001).
RN [18]
RP VARIANT TD THR-255, AND VARIANT ATHEROSCLEROSIS ASP-1611,
RX MEDLINE-21645894; PubMed-11785958;
RA Nishida Y., Hirano K., Tsukamoto K., Nagano M., Ikegami C., Roomp K.,
RA Ishihara M., Sakane N., Zhang Z., Tsujii K., Matsuyama A., Ohama T.,
RA Matsura F., Ishigami M., Sakai N., Hirooka H., Hattori H.,
RA Yamashita S., Matsuzawa Y.;
RT "Expression and functional analyses of novel mutations of ATP-binding
RT cassette transporter-1 in Japanese patients with high-density
RT lipoprotein deficiency.";
RL Biochem. Biophys. Res. Commun. 290:713-721(2002).
CC -1- FUNCTION: CAMP-DEPENDENT AND SUBSTRATE-SENSITIVE ANION
CC TRANSPORT. KEY GATEKEEPER INFLUENCING INTRACELLULAR CHOLESTEROL
CC TRANSPORT.
CC -1- TISSUE SPECIFICITY: WIDELY EXPRESSED, BUT MOST ABUNDANT IN
CC MACROPHAGES.
CC -1- DOMAIN: MULTIFUNCTIONAL POLYPEPTIDE WITH TWO HOMOLOGOUS HALVES,
CC EACH CONTAINING AN HYDROPHOBIC MEMBRANE-ANCHORING DOMAIN AND AN
CC ATP BINDING CASSETTE (ABC) DOMAIN.
CC -1- DISEASE: DEFECTS IN ABCA1 ARE A CAUSE OF HIGH DENSITY LIPOPROTEIN
CC DEFICIENCY TYPE I (HDLI), ALSO KNOWN AS TANGIER DISEASE (TD). TD
CC IS A RECESSIVE DISORDER CHARACTERIZED BY ABSENCE OF HIGH DENSITY
CC LIPOPROTEIN (HDL) CHOLESTEROL FROM PLASMA, HEPATOSPLENOMEGALY,
CC PERIPHERAL NEUROPATHY, AND FREQUENTLY PREMATURE CORONARY ARTERY
CC DISEASE (CAD).
CC -1- DISEASE: Defects in ABCA1 are a cause of high density lipoprotein



Query Match 99.8%; Score 11770; DB 1; Length 2261;
 Best Local Similarity 99.7%; Pred. No. 0;
 Matches 2255; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

1 MACWQPLRLMKNTFRRCOTCOLLEAVMPLFLLILISVRLSPYPEOHECHFPNKA 60
 1 MACWQPLRLMKNTFRRCOTCOLLEAVMPLFLLILISVRLSPYPEOHECHFPNKA 60
 61 MSAGTLPWVGIIICNANNPCFRYPTEGAPGVGNFNKSIYARLFSPARLLYSQDT 120
 61 MSAGTLPWVGIIICNANNPCFRYPTEGAPGVGNFNKSIYARLFSPARLLYSQDT 120
 121 SKMDRKVLRITLOQIKKSSNLKLODFLVNTEFSGLTYHNLSPKSTYDKMLRADVLH 180
 121 SKMDRKVLRITLOQIKKSSNLKLODFLVNTEFSGLTYHNLSPKSTYDKMLRADVLH 180
 181 KVFLOGYOLHLSLNGSKSEEMIOGDOEVELCGLPREKILAAERVLRSMIDLKPTL 240
 181 KVFLOGYOLHLSLNGSKSEEMIOGDOEVELCGLPREKILAAERVLRSMIDLKPTL 240
 241 RLINSTSPSPSKELAEATKTLHSLGTLAQLFMSRMSMDROEVMFLTNVSSSSSTQI 300
 241 RLINSTSPSPSKELAEATKTLHSLGTLAQLFMSRMSMDROEVMFLTNVSSSSSTQI 300
 301 YQAVSRIVCGHPEGGLIKSLNMYEDNNYKALFEGNGTEDEDAETFYDNSTTPYCNLDMK 360
 301 YQAVSRIVCGHPEGGLIKSLNMYEDNNYKALFEGNGTEDEDAETFYDNSTTPYCNLDMK 360
 361 NLESSPLRIIMKALKPLVYKIIYTPPTPATROYMAEVNTPFOELAFVHLEGMWELLS 420
 361 NLESSPLRIIMKALKPLVYKIIYTPPTPATROYMAEVNTPFOELAFVHLEGMWELLS 420
 421 PKIYTFMNSOBDLVRLDLSRDNDHFWEQDGLDWTADIVAFLLAKHPEDVOSSNGS 480
 421 PKIYTFMNSOBDLVRLDLSRDNDHFWEQDGLDWTADIVAFLLAKHPEDVOSSNGS 480
 481 VYTRAEAFENETNOAIRTSRMECVNLKLEPIATEVWLKXSHLDERFNAGIYFTG 540
 481 VYTRAEAFENETNOAIRTSRMECVNLKLEPIATEVWLKXSHLDERFNAGIYFTG 540
 541 ITPGSIELPHHVKKYKIRMDIDNVERTNKIKDGYMDPGRADPFEDMRYVWGCFAYLDDV 600
 541 ITPGSIELPHHVKKYKIRMDIDNVERTNKIKDGYMDPGRADPFEDMRYVWGCFAYLDDV 600
 601 EOLAIRVLTEGKTYGYMOOMPICYVDIPLFVNSRSMPLFMTLAMIYSVAVIIGIV 660
 601 EOLAIRVLTEGKTYGYMOOMPICYVDIPLFVNSRSMPLFMTLAMIYSVAVIIGIV 660
 661 YEKARLKETMRINGLDSILMFSMFISLILPLVSAGLLVILKIGNLPLPSDPVYFV 720
 661 YEKARLKETMRINGLDSILMFSMFISLILPLVSAGLLVILKIGNLPLPSDPVYFV 720
 721 FLSPFAVVTIIQCFILISTLFSRANLAAACGIIYFTLYLPVLCVAMQDYVGFLLKIFAS 780
 721 FLSPFAVVTIIQCFILISTLFSRANLAAACGIIYFTLYLPVLCVAMQDYVGFLLKIFAS 780
 781 LLSVAFGFGCEYALFEEDGIGYMNLPESPEYEDGFINLITISMLFDTFLYGVMTW 840
 781 LLSVAFGFGCEYALFEEDGIGYMNLPESPEYEDGFINLITISMLFDTFLYGVMTW 840
 841 YIEAVFGOYGIIPRPWFPCPKSWPGEBSDESHPSNOKRISEICMEDEPHLKIGVS 900
 841 YIEAVFGOYGIIPRPWFPCPKSWPGEBSDESHPSNOKRISEICMEDEPHLKIGVS 900
 901 IONLVKYYRGMKAVADGLALNFYEGQITSLGHNGAGKTTTMSILTLGPPTSGAYIL 960
 901 IONLVKYYRGMKAVADGLALNFYEGQITSLGHNGAGKTTTMSILTLGPPTSGAYIL 960
 961 GKDIRSEMSTIRONLGYCPOHNVLFQMLTYVEHIMFYARLKGISEKHVAKEMOMALDVG 1020
 961 GKDIRSEMSTIRONLGYCPOHNVLFQMLTYVEHIMFYARLKGISEKHVAKEMOMALDVG 1020

1021 LPSSKIKSTQSLSGQORLSTALAVGSSKVVIIIDEPPAGVDPPYSRGIWELLKTYRQ 1080
 1021 LPSSKIKSTQSLSGQORLSTALAVGSSKVVIIIDEPPAGVDPPYSRGIWELLKTYRQ 1080
 1081 GRTIILSTHMHDEADVGDRIAIISHGKLCCVSSSFLKNOLOETGYLTLYKKDVSLS 1140
 1081 GRTIILSTHMHDEADVGDRIAIISHGKLCCVSSSFLKNOLOETGYLTLYKKDVSLS 1140
 1141 SCRNSSTVSYLLKEDSVSSSSDAGIGSDHSDTLITIDVSAISNLRKHVSARLVEDI 1200
 1141 SCRNSSTVSYLLKEDSVSSSSDAGIGSDHSDTLITIDVSAISNLRKHVSARLVEDI 1200
 1201 GHELTLYLPEAKKEGAFVLFHEIDRLSDIGISSYGISSETLLEETFLKVAEESVDAE 1260
 1201 GHELTLYLPEAKKEGAFVLFHEIDRLSDIGISSYGISSETLLEETFLKVAEESVDAE 1260
 1261 TSDGTLPARNRRAFGDKOSCLRPTEDEADPNDSIDIPESRETELISGMDKGSYQK 1320
 1261 TSDGTLPARNRRAFGDKOSCLRPTEDEADPNDSIDIPESRETELISGMDKGSYQK 1320
 1321 GWKLTQOQFVALLMKRLILARRSRKGFPAQIYPAVFCIALVSLIYPPCKYPSLEQ 1380
 1321 GWKLTQOQFVALLMKRLILARRSRKGFPAQIYPAVFCIALVSLIYPPCKYPSLEQ 1380
 1381 PMWYNEQYTFVSNDAPEDTGTLLELNAITKDPGEGRKMGNGIPDTPOAGEEWTAP 1440
 1381 PMWYNEQYTFVSNDAPEDTGTLLELNAITKDPGEGRKMGNGIPDTPOAGEEWTAP 1440
 1441 VPOTIMDLFONGWMTQONPSPACOCSSDKIKKMLPVCPPGAGGLPPQKQNTADILQDL 1500
 1441 VPOTIMDLFONGWMTQONPSPACOCSSDKIKKMLPVCPPGAGGLPPQKQNTADILQDL 1500
 1501 TGRNISDYIKTYVOIILASLKXKIMVNEFRYGGESLGSVNTQALPPSOEVDAILKOMK 1560
 1501 TGRNISDYIKTYVOIILASLKXKIMVNEFRYGGESLGSVNTQALPPSOEVDAILKOMK 1560
 1561 HLKLANDSSADREPLNSLGRFMTGLDTRNNVKYWFENKKGHAISFLVNTNNAILRANLQK 1620
 1561 HLKLANDSSADREPLNSLGRFMTGLDTRNNVKYWFENKKGHAISFLVNTNNAILRANLQK 1620
 1621 GENPSHYGITAFNHLNLTROOLSEVALMTTSVDLVISICVIFAMSPVASEFVFLQDR 1680
 1621 GENPSHYGITAFNHLNLTROOLSEVALMTTSVDLVISICVIFAMSPVASEFVFLQDR 1680
 1681 VSKAKHLOFTSGVKPYIYMLSNFVMDKCYVVPATVYIIIFCFQOKSVYSTNLPLYAL 1740
 1681 VSKAKHLOFTSGVKPYIYMLSNFVMDKCYVVPATVYIIIFCFQOKSVYSTNLPLYAL 1740
 1741 LLLLYGMSITPLMPYASFEVKIPSTAYVVLTVSNLFTIGINSVATFVLELFTDNKLNNIN 1800
 1741 LLLLYGMSITPLMPYASFEVKIPSTAYVVLTVSNLFTIGINSVATFVLELFTDNKLNNIN 1800
 1801 DILKSVFLIFPHCLGRLIDMKYKNOAMADALERGENRFVPSLMDLVGRNLFMAAVEG 1860
 1801 DILKSVFLIFPHCLGRLIDMKYKNOAMADALERGENRFVPSLMDLVGRNLFMAAVEG 1860
 1861 VVEFLTIVLIQYREFIRPPVNAKLSPLNDEDEDEYRREORILDDGGONDILEIKELKI 1920
 1861 VVEFLTIVLIQYREFIRPPVNAKLSPLNDEDEDEYRREORILDDGGONDILEIKELKI 1920
 1921 YRRKRPAPVADRICVGIIPGECFGLLVGNAGKSTFKMLTGDYTVTRDAPLANKSISLN 1980
 1921 YRRKRPAPVADRICVGIIPGECFGLLVGNAGKSTFKMLTGDYTVTRDAPLANKSISLN 1980
 1981 IHEVHONMGYCOFPAITELLGHEHVEFFALLRCVPEKXVGEBAIRKLGIVKYEK 2040
 1981 IHEVHONMGYCOFPAITELLGHEHVEFFALLRCVPEKXVGEBAIRKLGIVKYEK 2040
 2041 YAGNTSGGNKRLSTAMALIGGPVYVFLDEPTTGMDPARPRFLMNCALSYVKEGSSVLT 2100
 2041 YAGNTSGGNKRLSTAMALIGGPVYVFLDEPTTGMDPARPRFLMNCALSYVKEGSSVLT 2100
 2101 SHSMECEALCTIRMAIMVNGRFRCLGSVOHLKNRFGDGYTIVVRIAGSNPDLKPVODFFG 2160


```

Db 2101 SHSMECEALCTRAIMAVNGRRCGLSVQHLKNRGDGYITLVIRAGSNPLKPVQDPFG 2160
Oy 2161 LAFPSVYKKEKRRMLQYQLPSSLSLARIRISLSQSKRRLLHEDYVSQTTLDQVYNF 2220
Db 2161 LAFPSVYKKEKRRMLQYQLPSSLSLARIRISLSQSKRRLLHEDYVSQTTLDQVYNF 2220
Oy 2221 AKDQSDDHKLKDLKRNQYVAVLRFSPLODEKVESYV 2261
Db 2221 AKDQSDDHKLKDLKRNQYVAVLRFSPLODEKVESYV 2261

RESULT 2
ID ABCL_MOUSE STANDARD: PRT: 2261 AA.
AC P41233.
DT 01-FEB-1995 (Rel. 31, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ATP-binding cassette, sub-family A, member 1 (ATP-binding cassette transporter 1) (ATP-binding cassette 1) (ABC-1).
GN ABCA1 OR ABC1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DKA/2; TISSUE=Macrophage;
RX MEDLINE=94375008; PubMed=8088782;
RA Luciani M.F., Denizot F., Savary S., Mattei M.-G., Chimi G.;
RT "Cloning of two novel ABC transporters mapping on human chromosome 9".
RL Genomics 21:150-159(1994).
RM [2]
RN SEQUENCE FROM N.A.
RP STRAIN=C57BL/6J;
RA Qiu Y., Cavelier L., Chiu S., Rubin E., Cheng J.-F.;
RT "Human and mouse ABCA1 comparative sequencing and transgenesis studies identify potential regulatory sequences."
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: CAMP-DEPENDENT AND SULFONYLUREA-SENSITIVE ANION TRANSPORTER. KEY GATEKEEPER INFLUENCING INTRACELLULAR CHOLESTEROL TRANSPORT (BY SIMILARITY).
CC -1- TISSUE SPECIFICITY: WIDELY EXPRESSED IN ADULT TISSUES. HIGHEST LEVELS ARE FOUND IN PREGNANT UTERUS AND UTERUS.
CC -1- DOMAIN: MULTIFUNCTIONAL POLYPEPTIDE WITH TWO HOMOLOGOUS HALVES, EACH CONTAINING AN HYDROPHOBIC MEMBRANE-ANCHORING DOMAIN AND AN ATP BINDING CASSETTE (ABC) DOMAIN.
CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. ABCA SUBFAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@sib-sib.ch).
CC
CC EMBL: X75926; CAA53530.1; ALT_INIT.
DR EMBL: AF287263; AAG39073.1; ALT_INIT.
DR MGI: 99607; Abcal.
DR InterPro: IPR003593; AAA_Arpase.
DR InterPro: IPR003439; ABC_transporter.
DR Pfam: PF00005; ABC_tran. 2.
DR ProDom: PD000006; ABC_transportr; 2.
DR SMART: SM00382; AAA; 1.
DR PROSITE: PS00211; ABC_TRANSPORTER; 1.
KW ATP-binding; Glycoprotein; Transmembrane; Transport.
FT TRANSMEM 26 42 POTENTIAL.
FT TRANSMEM 640 656 POTENTIAL.
FT TRANSMEM 690 706 POTENTIAL.
FT TRANSMEM 717 733 POTENTIAL.

```

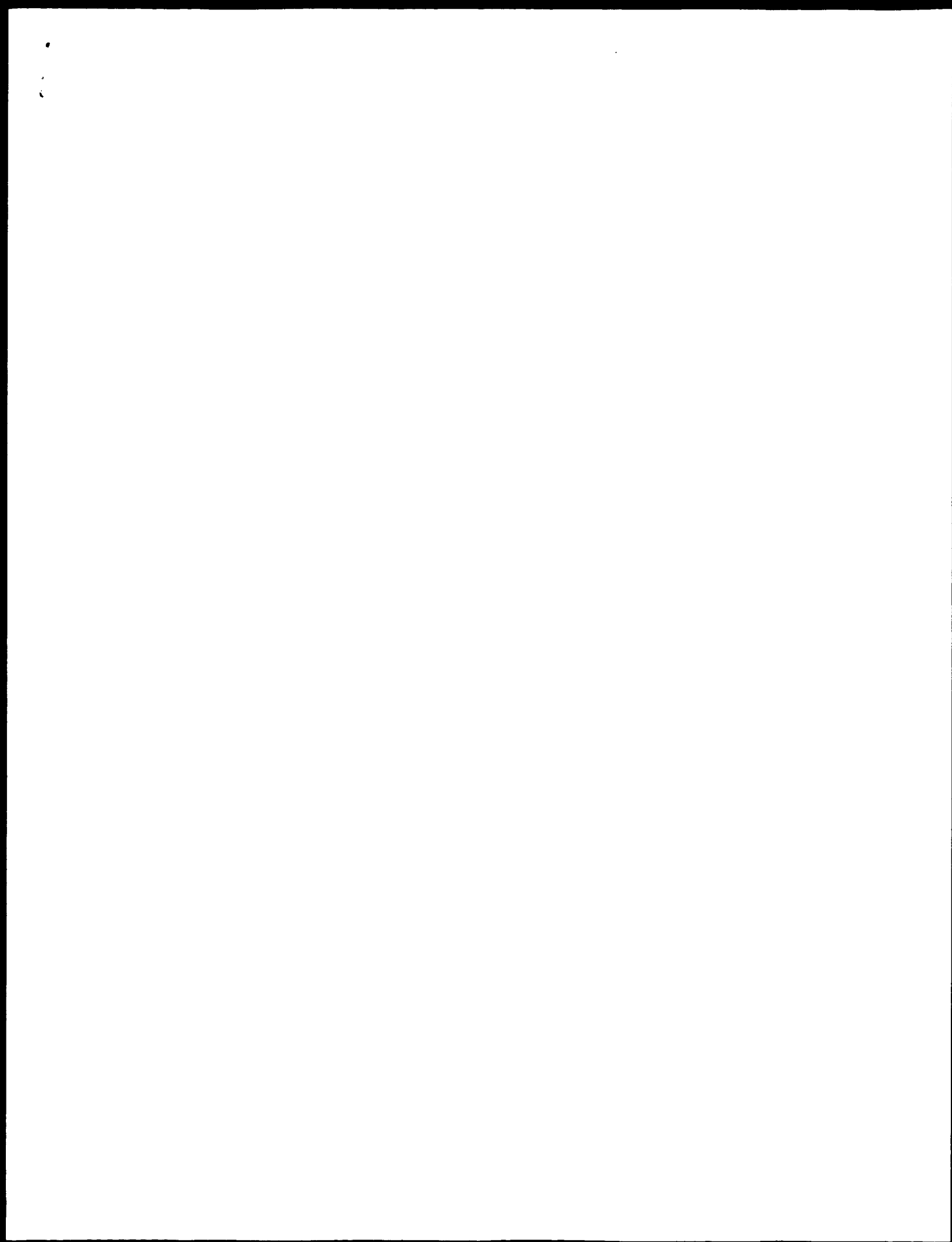
```

FT TRANSMEM 749 765 POTENTIAL.
FT TRANSMEM 771 787 POTENTIAL.
FT TRANSMEM 1041 1057 POTENTIAL.
FT TRANSMEM 1351 1367 POTENTIAL.
FT TRANSMEM 1661 1677 POTENTIAL.
FT TRANSMEM 1708 1724 POTENTIAL.
FT TRANSMEM 1737 1753 POTENTIAL.
FT TRANSMEM 1775 1791 POTENTIAL.
FT TRANSMEM 1854 1870 POTENTIAL.
FT NP_BIND 933 940 ATP (POTENTIAL).
FT NP_BIND 1946 1953 ATP (POTENTIAL).
FT CARBOHYD 14 14 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 98 98 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 151 151 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 161 161 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 196 196 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 244 244 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 292 292 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 337 337 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 349 349 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 400 400 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 478 478 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 489 489 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 521 521 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 820 820 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1144 1144 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1294 1294 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1453 1453 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1499 1499 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1504 1504 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1637 1637 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2044 2044 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2238 2238 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CONFLICT 1567 1568 MISSING (IN REF. 2).
FT CONFLICT 2024 2024 MISSING (IN REF. 2).
SQ SEQUENCE 2261 AA; 254011 MW; FAE62B21FDD09F9 CRC64;

Query Match 95.4%; Score 11256; DB 1; Length 2261;
Best Local Similarity 95.0%; Pred. No. 0;
Matches 2149; Conservative 55; Mismatches 57; Indels 0; Gaps 0;

Oy 1 MACPOLRLMLKMLNFRRRQTCOLLLEVAMPFLIFILISVRLSPYVQHCHEPPNKA 60
Db 1 MACPOLRLMLKMLNFRRRQTCOLLLEVAMPFLIFILISVRLSPYVQHCHEPPNKA 60
Oy 61 MPSAGTLPWVOGIIICANNNCFRRPPEGEAPGVGNFKSIVARLSDSARLLYSOKRT 120
Db 61 MPSAGTLPWVOGIIICANNNCFRRPPEGEAPGVGNFKSIVARLSDSARLLYSOKRT 120
Oy 121 SMKMRVRLFTLOQIKKSSNLKLODFLVNENFSGFLVNNLSLPKSTVIMRADYIIL 180
Db 121 SMKMRVRLFTLOQIKKSSNLKLODFLVNENFSGFLVNNLSLPKSTVIMRADYIIL 180
Oy 181 KVFLOGYOLHTLSJCNSSKSEEMITQDQVSELSGLPKKRLAAERVLRSNMDILKPL 240
Db 181 KVFLOGYOLHTLSJCNSSKSEEMITQDQVSELSGLPKKRLAAERVLRSNMDILKPL 240
Oy 181 KVFLOGYOLHTLSJCNSSKSEEMITQDQVSELSGLPKKRLAAERVLRSNMDILKPL 240
Db 181 KVFLOGYOLHTLSJCNSSKSEEMITQDQVSELSGLPKKRLAAERVLRSNMDILKPL 240
Oy 241 RTLSTSPFPEKELAEATKTLHSLGTLAQLSFMKWSMDRQEWMLTVNWNSSSSTOI 300
Db 241 RTLSTSPFPEKELAEATKTLHSLGTLAQLSFMKWSMDRQEWMLTVNWNSSSSTOI 300
Oy 301 YQAVSRIVCGHPREGGGLKIKSLNMYEDNNYKALPGGNGTEDEATFYDNSTTPYCNLIM 360
Db 301 YQAVSRIVCGHPREGGGLKIKSLNMYEDNNYKALPGGNGTEDEATFYDNSTTPYCNLIM 360
Oy 361 NLESSPLSRITWKALKPLLVKILYTPDTPATROYMAEVNKKTEPQLAVFDLREGMEELS 420
Db 361 NLESSPLSRITWKALKPLLVKILYTPDTPATROYMAEVNKKTEPQLAVFDLREGMEELS 420
Oy 421 PKITPFMSNGBMLVRLNLDSDNDHFWRQQLDGLDWTADYIAFLAKRPEDVQSSNGS 480
Db 421 PKITPFMSNGBMLVRLNLDSDNDHFWRQQLDGLDWTADYIAFLAKRPEDVQSSNGS 480

```



QY 241 RLINSTSPSPSKELAEATKTLHSLGILAELEFSMRKSWMDROEMFLTNVSSSSSTOI 300
 DB 241 RLINSTSPSPSKELAEATKTLHSLGILAELEFSMRKSWMDROEMFLTNVSSSSSTOI 300
 QY 301 YOAVSRIVCGHPEGGLKIKSLNWEEDNNYKALFEGNGSTEDDAETFYDNTTTPYCNDLMK 360
 DB 301 YOAVSRIVCGHPEGGLKIKSLNWEEDNNYKALFEGNGSTEDDAETFYDNTTTPYCNDLMK 360
 QY 361 NLESSPRLRIIMKALKPLLVCKIILYTPPTPATROYMAEVNTPQELAVFHLDEGMWELIS 420
 DB 361 NLESSPRLRIIMKALKPLLVCKIILYTPPTPATROYMAEVNTPQELAVFHLDEGMWELIS 420
 QY 421 PKIWMFEMENSOEMDLVRLMLDSRDNDHFEQOOLGDLMTAODIYAFLAKHEDVQSSNGS 480
 DB 421 PKIWMFEMENSOEMDLVRLMLDSRDNDHFEQOOLGDLMTAODIYAFLAKHEDVQSSNGS 480
 QY 481 VYTMREAEENEMOALRTISREMECVNLKLEPIATEVWLJNKSMLDERKFMAGIYFTG 540
 DB 481 VYTMREAEENEMOALRTISREMECVNLKLEPIATEVWLJNKSMLDERKFMAGIYFTG 540
 QY 541 ITPGSEILPHHVKYKIRMDIDNVERTNKIKDGYMDPGRADPFEDMRVWVGAFVLODV 600
 DB 541 ITPGSEILPHHVKYKIRMDIDNVERTNKIKDGYMDPGRADPFEDMRVWVGAFVLODV 600
 QY 601 EOAIRBVLTGTEKKTGYVMQMPYCYVDITFLRVMSRSMPLMTLANTYVAVIIGIY 660
 DB 601 EOAIRBVLTGTEKKTGYVMQMPYCYVDITFLRVMSRSMPLMTLANTYVAVIIGIY 660
 QY 661 YKKEARLEKEMIMGLDINSILMFSEISLIPPLVSAGLVILKLGMLLSPDSVYEV 720
 DB 661 YKKEARLEKEMIMGLDINSILMFSEISLIPPLVSAGLVILKLGMLLSPDSVYEV 720
 QY 721 FLVSFAVVTILLOCFILSTLFSRANLAACGGIYFTLYPLVLCVAMODYVGTFLKIFAS 780
 DB 721 FLVSFAVVTILLOCFILSTLFSRANLAACGGIYFTLYPLVLCVAMODYVGTFLKIFAS 780
 QY 781 LLSPAFGGCEYFALFEEGOGIYQWMDLFEPSVEDEGNLTTSISMLPOTFLYGVMTW 840
 DB 781 LLSPAFGGCEYFALFEEGOGIYQWMDLFEPSVEDEGNLTTSISMLPOTFLYGVMTW 840
 QY 841 YLEAVFPGGYGILPRMYPFCTKSYWFGESDEKSHPSGNOKRMSRICMEEPETHLKIGVS 900
 DB 841 YLEAVFPGGYGILPRMYPFCTKSYWFGESDEKSHPSGNOKRMSRICMEEPETHLKIGVS 900
 QY 901 TONLYKVVYRDGKXVAVDGLALNFYEGQITSPFGHAGAKTJTMSTLTLGFPPTSGTAYIL 960
 DB 901 TONLYKVVYRDGKXVAVDGLALNFYEGQITSPFGHAGAKTJTMSTLTLGFPPTSGTAYIL 960
 QY 961 GADISEMSTIKQNLGVCPOHNVLFDMLTVEHIMFYARLKGJSEKHAKEMQALDVG 1020
 DB 961 GADISEMSTIKQNLGVCPOHNVLFDMLTVEHIMFYARLKGJSEKHAKEMQALDVG 1020
 QY 1021 LPSKLSKTSQSLSGMOKKLSVALAFVGGSKVILDEPTAGVDPYSRRGJWELLKTYRQ 1080
 DB 1021 LPSKLSKTSQSLSGMOKKLSVALAFVGGSKVILDEPTAGVDPYSRRGJWELLKTYRQ 1080
 QY 1081 GFTIILSTHMHMEAYLIGRITAIISHGKICCVGSSLFLKNOLGXYTLTLVKKDDESSL 1140
 DB 1081 GFTIILSTHMHMEAYLIGRITAIISHGKICCVGSSLFLKNOLGXYTLTLVKKDDESSL 1140
 QY 1141 SCRNSSSTVYLKEDSVQSSSDAGLSDHESDPLTIDVSAISMLIKRHVSEALVEDI 1200
 DB 1141 SCRNSSSTVYLKEDSVQSSSDAGLSDHESDPLTIDVSAISMLIKRHVSEALVEDI 1200
 QY 1201 GHELTYYVLPYBAKBGAYELFHEITDRLISLGISSYGISETTLEIFLKYAESGVDAE 1260
 DB 1201 GHELTYYVLPYBAKBGAYELFHEITDRLISLGISSYGISETTLEIFLKYAESGVDAE 1260
 QY 1261 TSDGTLPARARNRARGDKOSCLRPTEDDAADPNDSDIDPESRETDLLSGMDKGSYVYK 1320
 DB 1261 TSDGTLPARARNRARGDKOSCLRPTEDDAADPNDSDIDPESRETDLLSGMDKGSYVYK 1320
 QY 1321 GMLTQOQFVALLMKRLILARRSRKGFPAQIVLPAVFCIALVFSLIIVPGKYPSELEQ 1380

DB 1321 GMLTQOQFVALLMKRLILARRSRKGFPAQIVLPAVFCIALVFSLIIVPGKYPSELEQ 1380
 QY 1381 PMWNEQYTFVSNDADEBDGTELLNALTKDPCFGRCMEGNPIPDTPQOAGEEEMTAP 1440
 DB 1381 PMWNEQYTFVSNDADEBDGTELLNALTKDPCFGRCMEGNPIPDTPQOAGEEEMTAP 1440
 QY 1441 VPOTIMDLFONGMWTQONSPACOCSSDKIKRMLPVCPPGAGGLPPPOKQNTADILDL 1500
 DB 1441 VPOTIMDLFONGMWTQONSPACOCSSDKIKRMLPVCPPGAGGLPPPOKQNTADILDL 1500
 QY 1501 TGRNIDSVLYKTYVOIILASLKNK IWNNEFRIGGSLVSNTOALPPSOEVNDAIKOMK 1560
 DB 1501 TGRNIDSVLYKTYVOIILASLKNK IWNNEFRIGGSLVSNTOALPPSOEVNDAIKOMK 1560
 QY 1561 HLKLAQSSADRLNSLGRMTGLDTRNNVYKWFNNKGHAISFLNVAIILANLOK 1620
 DB 1561 HLKLAQSSADRLNSLGRMTGLDTRNNVYKWFNNKGHAISFLNVAIILANLOK 1620
 QY 1621 GENPSHYGITAFAHNPILNTKQOLSEVALMTSVLVLSICVIFAMSFPASFEVLEIOER 1680
 DB 1621 GENPSHYGITAFAHNPILNTKQOLSEVALMTSVLVLSICVIFAMSFPASFEVLEIOER 1680
 QY 1681 VSKAKHLOFTISGVKPYIYWSLNFVWDMCNVVPATLVIIIFICFOOKSVSSSTNLPLVAL 1740
 DB 1681 VSKAKHLOFTISGVKPYIYWSLNFVWDMCNVVPATLVIIIFICFOOKSVSSSTNLPLVAL 1740
 QY 1741 LLLLGWSTIPLEMPASFEVKIPSTAYVLTSVNLFIGINGSVAFTLELFDNKLNNIN 1800
 DB 1741 LLLLGWSTIPLEMPASFEVKIPSTAYVLTSVNLFIGINGSVAFTLELFDNKLNNIN 1800
 QY 1801 DILKSVEFLRPHFCLOGRLIDMYKNOAMADALERGENRVSPLSMDLVGRULPMAVVG 1860
 DB 1801 DILKSVEFLRPHFCLOGRLIDMYKNOAMADALERGENRVSPLSMDLVGRULPMAVVG 1860
 QY 1861 VVEFLITVLQYRFFTRPRVNAKISPLNDEDEVDVREORILIDGGGONDIEIKELTKI 1920
 DB 1861 VVEFLITVLQYRFFTRPRVNAKISPLNDEDEVDVREORILIDGGGONDIEIKELTKI 1920
 QY 1921 YRRKRPAVDRICVGIIPGCEFGLLGVNAGKSTFKMLTGTTVYRGDAFLNRNLSIN 1980
 DB 1921 YRRKRPAVDRICVGIIPGCEFGLLGVNAGKSTFKMLTGTTVYRGDAFLNRNLSIN 1980
 QY 1981 IHEYHOMGVCPOFDATITELLTGREHVEFPALLRGVPEKEVGVMATRKIGLVKYGK 2040
 DB 1981 IHEYHOMGVCPOFDATITELLTGREHVEFPALLRGVPEKEVGVMATRKIGLVKYGK 2040
 QY 2041 YAGNYSGKNRKLSTAMALIGCPPVVEFLDEPTGMDPKARFLMNCALSVEKGRSVLT 2100
 DB 2041 YAGNYSGKNRKLSTAMALIGCPPVVEFLDEPTGMDPKARFLMNCALSVEKGRSVLT 2100
 QY 2101 SHSMECEALCTYMAIYNVNGRFCLGSVOHLKRRFEDGTYIYVRIAGSNPDLKVPQDFEG 2160
 DB 2101 SHSMECEALCTYMAIYNVNGRFCLGSVOHLKRRFEDGTYIYVRIAGSNPDLKVPQDFEG 2160
 QY 2161 LAEPGSLYKKEHNNMLOYPSSLSLARIFFSILSOSKRLHIEDYSVQOTLLDQVFVNF 2220
 DB 2161 LAEPGSLYKKEHNNMLOYPSSLSLARIFFSILSOSKRLHIEDYSVQOTLLDQVFVNF 2220
 QY 2221 AKQOSDDHLKDLSLKKNQTVVDVAVLTSFLQDEKVESTV 2261
 DB 2221 AKQOSDDHLKDLSLKKNQTVVDVAVLTSFLQDEKVESTV 2261

RESULT 2
 US-09-846-456-11
 : Sequence 11, Application US/09846456
 : Patent No. US20020146792A1
 : GENERAL INFORMATION:
 : APPLICANT: Rosier, Marie
 : APPLICANT: Prades, Catherine
 : APPLICANT: Lemoine, Cendriline
 : APPLICANT: Naudin, Laurent

APPLICANT: Deneffe, Patrice
APPLICANT: Duveger, Nicolas
APPLICANT: Brewer, Bryan
APPLICANT: Remaley, Alan
APPLICANT: Foto, Silvia
TITLE OF INVENTION: Regulatory Nucleic Acid for the ABC1 Gene, Molecules Modifying It
FILE REFERENCE: 3806.0505
CURRENT APPLICATION NUMBER: US/09/846.456
CURRENT FILING DATE: 2001-05-02
PRIORITY APPLICATION NUMBER: US 60/201,280
PRIORITY FILING DATE: 2000-05-02
NUMBER OF SEQ ID NOS: 20
SOFTWARE: Patentin version 3.0
SEQ ID NO 11
LENGTH: 2261
TYPE: PR
ORGANISM: Homo sapiens
US-09-846-456-11

Query Match 99.7%; Score 11767; DB 10; Length 2261;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 2254; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 MACMPQLRLMLKMLTERRRROTQQLLEVAMPFLFILILISRLSTPYEQHECHPNKA 60
DB 1 MACMPQLRLMLKMLTERRRROTQQLLEVAMPFLFILILISRLSTPYEQHECHPNKA 60
QY 61 MPAGTLPWVGITICNANNPCFRYPTEGAEAGVYGNFNKSTVARLFSDARLLYSOKDT 120
DB 61 MPAGTLPWVGITICNANNPCFRYPTEGAEAGVYGNFNKSTVARLFSDARLLYSOKDT 120
QY 121 SMKDMRVLTLOQIKKSSNMKLQDELVDNETSGFLYHNLSPKSTVVKMLRADYLH 180
DB 121 SMKDMRVLTLOQIKKSSNMKLQDELVDNETSGFLYHNLSPKSTVVKMLRADYLH 180
QY 121 SMKDMRVLTLOQIKKSSNMKLQDELVDNETSGFLYHNLSPKSTVVKMLRADYLH 180
DB 121 SMKDMRVLTLOQIKKSSNMKLQDELVDNETSGFLYHNLSPKSTVVKMLRADYLH 180
QY 181 KVFLOQVQLHLTSLCNSKSEEMIQLODOEYSELGCLPKKEKLAABVLRNSMDILKPL 240
DB 181 KVFLOQVQLHLTSLCNSKSEEMIQLODOEYSELGCLPKKEKLAABVLRNSMDILKPL 240
QY 241 RLTLSTSPFPKELAEATKTLHSLGTLAOELFMSRMSMDROVMFLTNVNSSSTOI 300
DB 241 RLTLSTSPFPKELAEATKTLHSLGTLAOELFMSRMSMDROVMFLTNVNSSSTOI 300
QY 301 YOAVSRIVCGHPEGGLKIKSLNMYEDNNYKALFGNGTEDEATFPYDNSTTPYCNLDM 360
DB 301 YOAVSRIVCGHPEGGLKIKSLNMYEDNNYKALFGNGTEDEATFPYDNSTTPYCNLDM 360
QY 361 NLESSPLSRIITWALKLPLVGLKILYPTDPATROVMAEVNKTFOELAVFHDLEGMEELS 420
DB 361 NLESSPLSRIITWALKLPLVGLKILYPTDPATROVMAEVNKTFOELAVFHDLEGMEELS 420
QY 421 PKITPMENSGEMDLVRLILDSRNDHFWEQQLDGLMTMODIYAFLAKHPEDVSSNGS 480
DB 421 PKITPMENSGEMDLVRLILDSRNDHFWEQQLDGLMTMODIYAFLAKHPEDVSSNGS 480
QY 481 VYTWREAFNETNOAIRTISRMECVNLKLEPIATEYWLINLKSMLLDERKFMAGIYFTG 540
DB 481 VYTWREAFNETNOAIRTISRMECVNLKLEPIATEYWLINLKSMLLDERKFMAGIYFTG 540
QY 541 ITTSGSIELPHVVKIKKIMQIDNVERTNKKIDGWDPCGRADPEEDMRYVMGFAYLDDV 600
DB 541 ITTSGSIELPHVVKIKKIMQIDNVERTNKKIDGWDPCGRADPEEDMRYVMGFAYLDDV 600
QY 601 EOAIIRVITGTEKTKGYVMQMPYCYVDLFLVMSRSMPLFMTLAMIYSVAATIKGIV 660
DB 601 EOAIIRVITGTEKTKGYVMQMPYCYVDLFLVMSRSMPLFMTLAMIYSVAATIKGIV 660
QY 661 YEKEARLEKTRINGLDNLSILMSFISLILPLVAGLVLVILKGNLLPYSDPSVVF 720
DB 661 YEKEARLEKTRINGLDNLSILMSFISLILPLVAGLVLVILKGNLLPYSDPSVVF 720
QY 721 FLVSFAVVTIIQCLISTLSRANLAACGIIYFTLYLPYVLCVAMQDYVGFTLKIFAS 780

DB 721 FLVSFAVVTIIQCLISTLSRANLAACGIIYFTLYLPYVLCVAMQDYVGFTLKIFAS 780
QY 781 LLSVAFGCGCEYFALFEBOGIGVQWMDNLFESPVEDEGNLTSSIMLFDFTLVGMT 840
DB 781 LLSVAFGCGCEYFALFEBOGIGVQWMDNLFESPVEDEGNLTSSIMLFDFTLVGMT 840
QY 841 YIEAVFPGGYGIPRPWYFPCTKSYWFGESDEKSHPGSNOKRMEISEEPTHLKGV 900
DB 841 YIEAVFPGGYGIPRPWYFPCTKSYWFGESDEKSHPGSNOKRMEISEEPTHLKGV 900
QY 901 IONLVVYVYRDGKVAVDGLALNFBEQITISFLGHGACGTTTMSLITGLFPPTSGTA 960
DB 901 IONLVVYVYRDGKVAVDGLALNFBEQITISFLGHGACGTTTMSLITGLFPPTSGTA 960
QY 961 GKDIREMSTIRONLGVCGQHNVLFDMLTVEEHIMFYARLKLSEKHVKAEMQALDVG 1020
DB 961 GKDIREMSTIRONLGVCGQHNVLFDMLTVEEHIMFYARLKLSEKHVKAEMQALDVG 1020
QY 1021 LPSSKLKSTQSLSGGMOKRLSVALAFVGSKVYLDEPTAGVDYPSRRCIWEILLKYR 1080
DB 1021 LPSSKLKSTQSLSGGMOKRLSVALAFVGSKVYLDEPTAGVDYPSRRCIWEILLKYR 1080
QY 1081 GRTIILSTHMDADVLGRIALISHGKLCVCGSSFLKNOIGTYLTLYKKDVSSLS 1140
DB 1081 GRTIILSTHMDADVLGRIALISHGKLCVCGSSFLKNOIGTYLTLYKKDVSSLS 1140
QY 1141 SCRNSSSTVYLYKKEDESVQSSDAGLSDHESDPTITIVSAISMLIRKHVEARLVED 1200
DB 1141 SCRNSSSTVYLYKKEDESVQSSDAGLSDHESDPTITIVSAISMLIRKHVEARLVED 1200
QY 1201 GHELTLYVLEAAGEAFVLEFHEIDRLSDLGISYGETTLEEIFLKAEEGVD 1260
DB 1201 GHELTLYVLEAAGEAFVLEFHEIDRLSDLGISYGETTLEEIFLKAEEGVD 1260
QY 1261 TSDGTLPARNRARAFDCKOSCLRPETEDADPNDSIDIPESREFDLSGMDKGSY 1320
DB 1261 TSDGTLPARNRARAFDCKOSCLRPETEDADPNDSIDIPESREFDLSGMDKGSY 1320
QY 1321 GMLTQOQFVALIMKRLLLARSRKGFQOIVLPAVFCIALVFSILVPPFKYSLE 1380
DB 1321 GMLTQOQFVALIMKRLLLARSRKGFQOIVLPAVFCIALVFSILVPPFKYSLE 1380
QY 1381 PMYNQYVFNVDNADPEDEGTLELNLALTKDPGFCRCMEGNPIPTPCQAGEEEMTAP 1440
DB 1381 PMYNQYVFNVDNADPEDEGTLELNLALTKDPGFCRCMEGNPIPTPCQAGEEEMTAP 1440
QY 1441 VPOTIMDLFONGMWTQNPSPACQCSSDKIKKMLPVCPPGAGLPPPOKONTADIL 1500
DB 1441 VPOTIMDLFONGMWTQNPSPACQCSSDKIKKMLPVCPPGAGLPPPOKONTADIL 1500
QY 1501 TGRNISDYLVTYVOTIIRKSLNKLIMVNEFRYGGFSLGVSNTQALPPSEVDADIK 1560
DB 1501 TGRNISDYLVTYVOTIIRKSLNKLIMVNEFRYGGFSLGVSNTQALPPSEVDADIK 1560
QY 1561 HLKLAADSSADREPLNSLGRFMTGLDTRNNVYKWFNNKGHAISSEFLVYNNALILRAN 1620
DB 1561 HLKLAADSSADREPLNSLGRFMTGLDTRNNVYKWFNNKGHAISSEFLVYNNALILRAN 1620
QY 1621 GENPSHYGTAFNHPNLNLKQOLESVALMTTSVDVLSICVIFAMSFVASFVFLIOER 1680
DB 1621 GENPSHYGTAFNHPNLNLKQOLESVALMTTSVDVLSICVIFAMSFVASFVFLIOER 1680
QY 1681 VSKAKHLQFISGVKPYTYWLSNFWDMCMYVVPATLVIIIFICFOOKSVSSSTNLPL 1740
DB 1681 VSKAKHLQFISGVKPYTYWLSNFWDMCMYVVPATLVIIIFICFOOKSVSSSTNLPL 1740
QY 1741 LLLLYGMSITPLMAPASFVKIPSTAYVVLTVSNLFTIGNSVATFVLELFTDNKLNIN 1800
DB 1741 LLLLYGMSITPLMAPASFVKIPSTAYVVLTVSNLFTIGNSVATFVLELFTDNKLNIN 1800
QY 1801 DILKSVFLIFPHECLGRLIDWKKNOAMADALERGENFVSPLSMDLVGRNLPMAVBC 1860

Db 1801 DILKSVFLIFPHFLCGLIDMVKNOAMADALEREGERNFVPSLWDLVGRNLFAMAVEG 1860
 QY 1861 VVEFLITVLIOYRFFIRPRPVNAKLSPLNDEDEDREROHILDDGGONDLLEIKELTKI 1920
 Db 1861 VVEFLITVLIOYRFFIRPRPVNAKLSPLNDEDEDREROHILDDGGONDLLEIKELTKI 1920
 QY 1921 YRRKRRPAVDRIICVGIPEGCEGGLGVNAGKSSFTKMLTGDITVTRODAFLNKSILSN 1980
 Db 1921 YRRKRRPAVDRIICVGIPEGCEGGLGVNAGKSSFTKMLTGDITVTRODAFLNKSILSN 1980
 QY 1981 IHEVONNNGYCPQOPATIELLIGREHVEFPALLRGVPRKEVKGKGEAMIRKLGLVKYGEK 2040
 Db 1981 IHEVONNNGYCPQOPATIELLIGREHVEFPALLRGVPRKEVKGKGEAMIRKLGLVKYGEK 2040
 QY 2041 YAGNSGKNRKRKLSTAMALIGGPPVFLDEPTTGMDPKARFELNMCALSVYKEGRSVLT 2100
 Db 2041 YAGNSGKNRKRKLSTAMALIGGPPVFLDEPTTGMDPKARFELNMCALSVYKEGRSVLT 2100
 QY 2101 SISMECECALCTIRMAIMVNGRFRCLGSVOHLKNRFGDGTIVVRLAGSNPDLKPYODEFG 2160
 Db 2101 SISMECECALCTIRMAIMVNGRFRCLGSVOHLKNRFGDGTIVVRLAGSNPDLKPYODEFG 2160
 QY 2161 IAFPSVYKEKHRLMLOYQLPSSLSLARIFSILSOSKRLHIEDYSQOTLDQVYVF 2220
 Db 2161 IAFPSVYKEKHRLMLOYQLPSSLSLARIFSILSOSKRLHIEDYSQOTLDQVYVF 2220
 QY 2221 AKDOSDDHLDKLSLHKNQTVVDAVLTSLQDEKVKESYV 2261
 Db 2221 AKDOSDDHLDKLSLHKNQTVVDAVLTSLQDEKVKESYV 2261

RESULT 3

US-09-995-542-9
 : Sequence 9, Application US/09995542
 : Patent No. US20020127647A1
 : GENERAL INFORMATION:
 : APPLICANT: Shutter, John
 : APPLICANT: Ullas, Ilaari
 : TITLE OF INVENTION: ATP-Binding Cassette Transporter-Like Molecules and
 : TITLE OF INVENTION: Uses Thereof
 : FILE REFERENCE: 00-658-A
 : CURRENT APPLICATION NUMBER: US/09/995, 542
 : CURRENT FILING DATE: 2001-11-28
 : PRIOR APPLICATION NUMBER: 60/253, 520
 : PRIOR FILING DATE: 2000-11-28
 : NUMBER OF SEQ ID NOS: 24
 : SOFTWARE: PatentIn Ver. 2.0
 : SEQ ID NO 9
 : LENGTH: 2201
 : TYPE: PRT
 : ORGANISM: Mus musculus
 : FEATURE:
 : NAME/KEY: UNSURE
 : LOCATION: (115)
 : OTHER INFORMATION: amino acid at this position is unknown
 US-09-995-542-9

Query Match 92.6% Score 10922; DB 10; Length 2201;
 Best Local Similarity 94.9%; Pred. No. 0;
 Matches 2089; Conservative 55; Mismatches 57; Indels 0; Gaps 0;

QY 61 MPSAGTLFWOGIICNANPCFRYPPTGCEAPGVGNFNKSIIVARLFSDARLLIYSOKDT 120
 Db 1 MPSAGTLFWOGIICNANPCFRYPPTGCEAPGVGNFNKSIIVARLFSDARLLIYSOKDT 60
 QY 121 SMKDMRKVLRLLOQIKKSSNLKLODFLVNMTSGFLYHNLSPKSTVDKMLRADIYILH 180
 Db 121 SMKDMRKVLRLLOQIKKSSNLKLODFLVNMTSGFLYHNLSPKSTVDKMLRADIYILH 120
 QY 181 KVFLOGYOLHTLSLNGSKSEEMIOLGDOVESEICGLPKREKLAARVLSNMOLIPIL 240
 Db 181 KVFLOGYOLHTLSLNGSKSEEMIOLGDOVESEICGLPKREKLAARVLSNMOLIPIL 240

Db 121 KVFLOGYOLHTLSLNGSKSEEMIOLGDOVESEICGLPKREKLAARVLSNMOLIPIL 180

QY 241 RTINSTSPPEKELAEATKTLTLLSLCTLAOELFSMRSMDSMDROEVMFLTNVSSSSSTOI 300
 Db 181 TKLNSTSHLPQHLAETATYLLDLSGLAOELFSMRSMDSMDROEVMFLTNVSSSSSTOI 240
 QY 301 YQAVSRIVCGHPGGGLKIKSLNWEEDNNYKALFGSGNGTEDEAETFYDNSTTPCYNDLAK 360
 Db 241 YQAVSRIVCGHPGGGLKIKSLNWEEDNNYKALFGSGNGTEDEAETFYDNSTTPCYNDLAK 300
 QY 361 NLESSPISRIYKALPPLVYKILYTPDDPATROVAAEYKRFQELAVHDLEGMEELS 420
 Db 301 NLESSPISRIYKALPPLVYKILYTPDDPATROVAAEYKRFQELAVHDLEGMEELS 360
 QY 421 PKITWEMENSQEMDLVRMLDSRDNDHEWEOQLDGLMTADQIVAFKAPREDVQSSMGS 480
 Db 361 PKITWEMENSQEMDLVRMLDSRDNDHEWEOQLDGLMTADQIVAFKAPREDVQSSMGS 420
 QY 481 VYTMREAFNETNOAIFTTISRFMECVNLKLEPIPEVRLINKSMELDERKRFWAGIVFG 540
 Db 421 VYTMREAFNETNOAIFTTISRFMECVNLKLEPIPEVRLINKSMELDERKRFWAGIVFG 480
 QY 541 ITPGSIELPHHYKKTIRMDIDNVERTNKTIDGYWDPGRADPEFEDMRVYWGGAFLQDQV 600
 Db 481 ITPGSIELPHHYKKTIRMDIDNVERTNKTIDGYWDPGRADPEFEDMRVYWGGAFLQDQV 540
 QY 601 EOAIIIVLTGTEKKTGVYQOMPYPCYVDDIPLRWASRMPLEPTLAMYSAVVIKGIY 660
 Db 541 EOAIIIVLTGTEKKTGVYQOMPYPCYVDDIPLRWASRMPLEPTLAMYSAVVIKGIY 600
 QY 661 YEKEARKETEMRIMGDINSTLWFMSFISLIPLASAGLVIYILKGNLPLPDSVYFV 720
 Db 601 YEKEARKETEMRIMGDINSTLWFMSFISLIPLASAGLVIYILKGNLPLPDSVYFV 660
 QY 721 FLTSFAVVTLLQCELLSTLFSRANLAACGGIYFTLYLVYLCVAMODYVGTPLKIPAS 780
 Db 661 FLTSFAVVTLLQCELLSTLFSRANLAACGGIYFTLYLVYLCVAMODYVGTPLKIPAS 720
 QY 781 LSPVAFGCECFEALFEEOGTGVQMDNLFESPVEDGPNLTSTISMLPDTFLGVMTW 840
 Db 721 LSPVAFGCECFEALFEEOGTGVQMDNLFESPVEDGPNLTSTISMLPDTFLGVMTW 780
 QY 841 YIEAVEPGOYGIIPRWYFPCTKSYWPEESDEKSHPGSNOKBMSSEICEEPEETHLKLVS 900
 Db 781 YIEAVEPGOYGIIPRWYFPCTKSYWPEESDEKSHPGSNOKBMSSEICEEPEETHLKLVS 840
 QY 901 IONLVYVYRDGMKVAVDGLALNFYEQITSPFIGNHAGKTTMSTLTGLPPTSGTAYIL 960
 Db 841 IONLVYVYRDGMKVAVDGLALNFYEQITSPFIGNHAGKTTMSTLTGLPPTSGTAYIL 900
 QY 961 GKDIREMSTIRONLGVCPQHNVLPFMTLVEEHIMFYARLKGISEKHVYAEKQMAIDVG 1020
 Db 901 GKDIREMSTIRONLGVCPQHNVLPFMTLVEEHIMFYARLKGISEKHVYAEKQMAIDVG 960
 QY 1021 LPSSKLKSTKSQSLSGMOKRLSVLAFVGGSKVYLIDEPTAGVDYPSRRGIWELLKTRQ 1080
 Db 961 LPSSKLKSTKSQSLSGMOKRLSVLAFVGGSKVYLIDEPTAGVDYPSRRGIWELLKTRQ 1020
 QY 1081 GRTIILSTHMPADVLGRIALISHGKLCVGGSSFLKNQIGTGYLTLVKKDVESSLS 1140
 Db 1021 GRTIILSTHMPADVLGRIALISHGKLCVGGSSFLKNQIGTGYLTLVKKDVESSLS 1080
 QY 1141 SCRNSSSTVYLLKEDSVSSQSSDAGLSDHESDPLTIDVSAISNLIRKHSEARLAVDI 1200
 Db 1081 SCRNSSSTVYLLKEDSVSSQSSDAGLSDHESDPLTIDVSAISNLIRKHSEARLAVDI 1140
 QY 1201 GHELTIVLYPEAKKAGAFELFHEIDRLSDGISSYGISETTLEIFLKAEEBGGVAE 1260
 Db 1141 GHELTIVLYPEAKKAGAFELFHEIDRLSDGISSYGISETTLEIFLKAEEBGGVAE 1200
 QY 1261 TSDGTLPARRRNRAFGDKQSLRPFTEDDADPNOSDDIPESRETDLLSGMGKGSYOVK 1320
 Db 1201 TSDGTLPARRRNRAFGDKQSLRPFTEDDADPNOSDDIPESRETDLLSGMGKGSYOVK 1260
 QY 1321 GWKLTQOQFVALLMKRLLIARRSRKGFQAQIVLPAVFCIALVFSIIVPPFGKYPSELQ 1380

```

1261 GAKLTQOQFAALLMKRLILARRSRKGFADIVLPAVFCALVFSILVPEFGYPSLEIQ 1320
1381 PMWNEQYTVSNDAPEDDTLTLLNALTKDPGFCRCMGNDIPDTPCQAGEENTAP 1440
1321 PMWNEQYTVSNDAPEDDTLTLLNALTKDPGFCRCMGNDIPDTPCQAGEENTAP 1380
1441 VPOTIMDFONGWMTQNPSPACQSSDKIKKMLPVCPPGAGLPPOROKONTADILQDL 1500
1381 VPOSIVDLFONGWMTQNPSPACQSSDKIKKMLPVCPPGAGLPPOROKONTADILQDL 1440
1501 TGRNISDYLVKTYVQITIAKSLAKKIWNFRYGFSLGVSNTQALPPSQEVNDAIKQMK 1560
1441 TGRNISDYLVKTYVQITIAKSLAKKIWNFRYGFSLGVSNTQALPPSQEVNDAIKQMK 1500
1561 HLKLANDSSADREPLNSLGRMTGLDTRNNKYVFNKGMHAIISFLVNNATIRANLQK 1620
1501 LKLTNDTSADREPLNSLGRMTGLDTRNNKYVFNKGMHAIISFLVNNATIRANLQK 1560
1621 GENPSHGITAFAHNPPLNLTKQOLSEVALMTTSVDLVLSICVIFAMSPVASFVFLQER 1680
1561 GENPSHGITAFAHNPPLNLTKQOLSEVALMTTSVDLVLSICVIFAMSPVASFVFLQER 1620
1681 VSKAKHLQFISGVKPYIYHLSNFMCMCNVVPATLVIIIFICFOOKSYVSSINLPLAL 1740
1621 VSKAKHLQFISGVKPYIYHLSNFMCMCNVVPATLVIIIFICFOOKSYVSSINLPLAL 1680
1741 LLLLYGMSITPLMYPASFEKIPSTAYVVLTSYNLFIGINGSVATVLELFDNKNLNIN 1800
1681 LLLLYGMSITPLMYPASFEKIPSTAYVVLTSYNLFIGINGSVATVLELFDNKNLNIN 1740
1801 DILKSVPLFPHFCGLGRGLDMYKNQAMADALEFERGENFRVSPULSDVGRNLFAMAVEG 1860
1741 DILKSVPLFPHFCGLGRGLDMYKNQAMADALEFERGENFRVSPULSDVGRNLFAMAVEG 1800
1861 VVEFLITVLIOYFEFTRPRPVNAKLSPLNDEDEDEVARREQRILIDGGQNDILEIKELTKI 1920
1801 VVEFLITVLIOYFEFTRPRPVNAKLSPLNDEDEDEVARREQRILIDGGQNDILEIKELTKI 1860
1921 YRRKRRPAVDRIQVGIPECEFCGLGVNGAGKSTFKMLTGPTTVYRGDAFLNKNSTLSN 1980
1861 YRRKRRPAVDRIQVGIPECEFCGLGVNGAGKSTFKMLTGPTTVYRGDAFLNKNSTLSN 1920
1981 IHEVHONMGYCOFODAITELLTGREHVEFFALLRGVPEKEVGKGEWATRLKGLVYKGR 2040
1921 IHEVHONMGYCOFODAITELLTGREHVEFFALLRGVPEKEVGKGEWATRLKGLVYKGR 1980
2041 YAGNYSGGNKRKRLSTAMALIGGPPVFLDEPTTGMDPKARRFLMNCALSVKEGSRVLT 2100
1981 YAGNYSGGNKRKRLSTAMALIGGPPVFLDEPTTGMDPKARRFLMNCALSVKEGSRVLT 2040
2101 SHSMECEALCTEMALIMVNGRFCLGSVOHLNRFEGDYTIIVRIAGSPDLKPVODFFG 2160
2041 SHSMECEALCTEMALIMVNGRFCLGSVOHLNRFEGDYTIIVRIAGSPDLKPVODFFG 2100
2161 LARPGSVLEKRRHMLQYOLPSSLSLARIIFSLSSOKRLHIEDYVSQOTLDOVFVNF 2220
2101 LARPGSVLEKRRHMLQYOLPSSLSLARIIFSLSSOKRLHIEDYVSQOTLDOVFVNF 2160
2221 AKDOSDDHLKDLSLHKNQTVVDAVLTSLFLODEKESYV 2261
2161 AKDOSDDHLKDLSLHKNQTVVDAVLTSLFLODEKESYV 2221

```

```

; FILE REFERENCE: 00-658-A
; CURRENT APPLICATION NUMBER: US/09/995,542
; CURRENT FILING DATE: 2001-11-28
; PRIOR APPLICATION NUMBER: 60/253,520
; PRIOR FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 2310
; TYPE: prt
; ORGANISM: Mus musculus
; US-09-995-542-10

Query Match          49.9%; Score 5888; DB 10; Length 2310;
Best Local Similarity 50.1%; Pred. No. 0;
Matches 1168; Conservative 362; Mismatches 624; Indels 178; Gaps 27;

6 QRLMLKMLNLFRRRQTCOLLLEVAMPILFLLISVRLSYPEYQHCHEPKNKAMPSAG 65
6 QRLMLKMLNLFRRRQTCOLLLEVAMPILFLLISVRLSYPEYQHCHEPKNKAMPSAG 65
6 TLPMVQIICNANNCFFRPTTPEAGPYVGNPNKSIYARLFSDARRLLLSQKDTSMKDM 125
6 TLPMVQIICNANNCFFRPTTPEAGPYVGNPNKSIYARLFSDARRLLLSQKDTSMKDM 125
6 LPLWLGIFCNMNNPCFONPTPGESPTVSNYNNSLARVYRDFQELFMDTPEVQHLGVY 125
6 LPLWLGIFCNMNNPCFONPTPGESPTVSNYNNSLARVYRDFQELFMDTPEVQHLGVY 125
6 KVLRTLOOI-----KKSSNLKLODFLDNENPFGFLNHLSPSTVDKMLRAD 176
6 KVLRTLOOI-----KKSSNLKLODFLDNENPFGFLNHLSPSTVDKMLRAD 176
6 MAELFTLSQPMFDLTHTHERFAGRLQIRDLKQDEALTLFLMRNIGSDVAHLVNSQ 185
6 MAELFTLSQPMFDLTHTHERFAGRLQIRDLKQDEALTLFLMRNIGSDVAHLVNSQ 185
6 VVEFGVAGVDPDELTDIACSEALLQRFILFSQRRGAQTVADALCPLSQYTLQMIEDTLY 245
6 VVEFGVAGVDPDELTDIACSEALLQRFILFSQRRGAQTVADALCPLSQYTLQMIEDTLY 245
6 SNMDLTKILTLNSTSPFSPKELAEATKTLHSLGTL--AQELFSKRW---SDMR- 282
6 SNMDLTKILTLNSTSPFSPKELAEATKTLHSLGTL--AQELFSKRW---SDMR- 282
6 ADVDFEK-----LFHVLPTLSDSSSGGIMLRWGCILSDLSP 282
6 ADVDFEK-----LFHVLPTLSDSSSGGIMLRWGCILSDLSP 282
6 -----QEVWF-----LTNVSSSSSTQIYOAVSRIYVCGPEGGGLKIKSLMNEYD 327
6 -----QEVWF-----LTNVSSSSSTQIYOAVSRIYVCGPEGGGLKIKSLMNEYD 327
6 RMQKTIHPSVQDILLWVSRPLONGGPTFTQMSIISDLGCIPEGGGSRFSEFNWYED 342
6 RMQKTIHPSVQDILLWVSRPLONGGPTFTQMSIISDLGCIPEGGGSRFSEFNWYED 342
6 NNYKALFPGNGCTEEDAEFTYDNTSTPYCNDLMKNLESSPLSRIITWAKLPLVKGILYTP 387
6 NNYKALFPGNGCTEEDAEFTYDNTSTPYCNDLMKNLESSPLSRIITWAKLPLVKGILYTP 387
6 NNYKALFIDSTRKPDAYSYDKRTTSPCNSLIQSLESNPLTKIAMRAKPLMKGILFTF 402
6 NNYKALFIDSTRKPDAYSYDKRTTSPCNSLIQSLESNPLTKIAMRAKPLMKGILFTF 402
6 DTPAROVMAENVNTPFOCLAFHOLEGMELSPKTYTFMNSQEBMLVYKMLDSRNDH 447
6 DTPAROVMAENVNTPFOCLAFHOLEGMELSPKTYTFMNSQEBMLVYKMLDSRNDH 447
6 DSPAARRIKMANNSTFEELDVVRKLVKAMEVGPDIWFFEKSTQMTVIRDTLOHPYKD 462
6 DSPAARRIKMANNSTFEELDVVRKLVKAMEVGPDIWFFEKSTQMTVIRDTLOHPYKD 462
6 FWEQDLDLMDTAODIVAFLEKHPEDVQSSNGSVYTWREAFNFTNOALRTISPMECVNL 507
6 FWEQDLDLMDTAODIVAFLEKHPEDVQSSNGSVYTWREAFNFTNOALRTISPMECVNL 507
6 FLNRGLGEGITTEAVLNFSNGPOEKQADMTSFDKMDIFNITDRILRANLOYECLVL 522
6 FLNRGLGEGITTEAVLNFSNGPOEKQADMTSFDKMDIFNITDRILRANLOYECLVL 522
6 NKEPIATEVWLKNSMELDERKFWAGIVFTGTPSIELPHVHKYKIRMDIDNVERTN 567
6 NKEPIATEVWLKNSMELDERKFWAGIVFTGTPSIELPHVHKYKIRMDIDNVERTN 567
6 KIKGYPMPGRAPPEDEMRVWNGGFAYLDVVOALIIIRVLTGTEKTYGYMOQMPPCY 627
6 KIKGYPMPGRAPPEDEMRVWNGGFAYLDVVOALIIIRVLTGTEKTYGYMOQMPPCY 627
6 KIKDRYVDSGRADVEDEFRYIKGFAVLDQVBOGIVKQOMAEPPIGVYLOMPPCF 642
6 KIKDRYVDSGRADVEDEFRYIKGFAVLDQVBOGIVKQOMAEPPIGVYLOMPPCF 642
6 VDIIFLRVMSRMBPLFMTLWAVISVAVITKIGIYKEARLKETIRINGLINSILMF5MT 687
6 VDIIFLRVMSRMBPLFMTLWAVISVAVITKIGIYKEARLKETIRINGLINSILMF5MT 687
6 VDSFMITLNRCPILPVLWAMISVSVMTKIVALEKELRLKETLKNQGVANAVIWMCFWL 702
6 VDSFMITLNRCPILPVLWAMISVSVMTKIVALEKELRLKETLKNQGVANAVIWMCFWL 702
6 SSILPLVYAGLAVYVILKGLNLPYSDPSVYFVLSFAVAVTIIQGLISTLFSRANLAA 747
6 SSILPLVYAGLAVYVILKGLNLPYSDPSVYFVLSFAVAVTIIQGLISTLFSRANLAA 747
6 DSFISMLISFLLTLFIMHGRILHAYSDPFIILFLAFATATIMOSFLSTLFSKASIAA 762
6 DSFISMLISFLLTLFIMHGRILHAYSDPFIILFLAFATATIMOSFLSTLFSKASIAA 762
6 ACGGIYFTLVPLVLCVAMODVYGFLTKIFASILSVAGFGCEYALFEEDGIGQWMD 807
6 ACGGIYFTLVPLVLCVAMODVYGFLTKIFASILSVAGFGCEYALFEEDGIGQWMD 807
6 ACGGIYFTLVPLVLCVAMODVYGFLTKIFASILSVAGFGCEYALFEEDGIGQWMD 822
6 ACGGIYFTLVPLVLCVAMODVYGFLTKIFASILSVAGFGCEYALFEEDGIGQWMD 822

```

```

QY 808 NLESPVEEDGPNLTJTSIMMLFDFLYGVMTYEAEPGQYIIPRWYPCOTSYWFG 867
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 823 NICKSPLEGEDEFSLMKMKMLDALYGLLAWYLDQVPPGDPYGLPWPYFLLOSXYWG 882
QY 868 -----EESDEKSHPGSNOKRMSSEICOMEPEPHILKIGVSIOMLVKXY 908
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 883 GEGCSTREERALEKTEPLETEEMEDPEHP---EGMDSFFERLEPGLVGVGVCKMLKVP 938
QY 909 RDGMKAVADGIALNFEYGOITSFELHNGAGKTTTMSILGLFPPTSGTAYILGKDIRSEM 968
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 939 EPSGRRAVDRLNTFEYENQITAFELHNGAGKTTTMSILGLFPPTSGTAYILGKDIRSEM 998
QY 969 STTRONLVCPOHNVLPDMITVEEHITWETARLKLGSEKHYVAKEMOMALDVLGPPSKLS 1028
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 999 DVVROSLGCMPOHNIIFHNLITVAEHITLIFYAOLKGRSWEAOLEMEMLEDOTGL-HHKRNE 1057
QY 1029 KTSOLSGOMOKLSVALAFVGGSKVILDEPAGVDPYRGRGWELELLYRGGRTTIST 1088
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1058 EADOLSGOMOKLSVALAFVGGSKVILDEPAGVDPYRGRGWELELLYRGGRTTIST 1117
QY 1089 HHMDEADVLGDRITAIISHGKLCVSSLEFLKNQDGTGYLLTVK--KDYES-----SL5 1140
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1118 HHMDEADVLGDRITAIISHGKLCVSSLEFLKNQDGTGYLLTVK--KDYES-----SL5 1177
QY 1141 SCNNSSSYVYLKEDSVSSDAGLGDHEDTLITIVSASNLIRKHVSEARLYEDI 1200
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1178 SCTSKGSTRCPYRVEIT-----EEQVLDGVOQLMDLVYHHVPEAKLVECI 1225
QY 1201 GHEITVLPYEAKEGAFVLEPHIEDRLSDIGISSYSETTLEIFELKVAEESGVDAE 1260
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1226 GQELIFLPPKNNKORAYASLFLRELETLADJLSSFGISDIPLEIFLKVIEDNAGSM 1285
QY 1261 TSPGTLPARNNRAFGKOSC-----LRPTE-----DDAADPDSIDIPESRETDLL 1308
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1286 FVGG---AQQKREOAGLRHPCASPTKEROYAOAPHTCSPGOVDPKQSPSEPED--- 1338
QY 1309 SGMDGKSYOVKMKLTQOQFVALLMKRLLIARSRRKGFACIIVDPAVVCIALVLSLV 1368
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1339 -----PEVNTGARLILQVQALLVKRHHITRSKRDOVAQVILPATVFLALMSTIV 1393
QY 1369 PPGKYPSELEQPMWTNEOYTFVSNDAPEDTGTLELLMALTYKDPGFGTROMECNPIPDPT 1428
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1394 PPGFEPALTLHPWYCHOYTFESMDEPNNEHLEVLADVLNPPCGNRCLEKEMWLEXP 1453
QY 1429 CQAGEEEMTAPYPOITMDLFONGNMTMONPSPACOCSSDKIKMLPVCPRGAGIIPPO 1488
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1454 C-INATISWKTSPVSPNTTLHFQOKMTAAHPSPSCSTREKLTMLPECPREGAGILPPO 1512
QY 1489 RKONTADILDLGRNISDVLKTVYVOIYIAKSLKNKIWNEEFYGGFSLGVSTQALPPS 1548
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1513 RTGRSTEVLODLNRTNSDVLKTYPALIRSSLSKFWNEQKYGISIG-GKLPALPIS 1571
QY 1549 QEVNDALIKOKKHLKLAKOSSADRFNLISGR-----FMTGLDTRNN 1589
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1572 GE-----ALVGFPLSGIGOMNNAVSGPVTREASKEMLDPLKHELTEDN 1613
QY 1590 VKYMNENKGMHATSSFLNVTNNALIRANLOKGNPSHYGTANRHLNLTQOOLSVALLM 1649
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1614 IKWPNFKGHALVSLFNVAHNAITLRASLPRDRPBEYGTITVISOPLNLTKEULSDITVL 1673
QY 1650 TTSVDLVLSICVIFAMSEVPASFVEILDERVSKAKHLOFISGVKPVIIWLSNFWDMCN 1709
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1674 TTSVDAVVALCVIFAMSEVPASFVEILDERVSKAKHLOFISGVKPVIIWLSNFWDMCN 1733
QY 1710 YVPVATLVITIFCFOOKSVSTNLPVIALLLILGWSITPLMTYASVFKIPSTAYV 1769
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1734 YAVSAGLVVIGFQOKKATSPDNLPALVSLMLYGMVAIIPMYAPASLFEVPSATYVA 1793
QY 1770 LTVSNLPTIGNSVAPFVLELFTDNK-LNNINDILKSVFLPFPHPCIGLIDIMVNOQM 1828
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1794 LSCANLPTIGNSAITFVLELFEENKRLIRFNMLKRLIVFPHFLGKGLDLALISQAV 1853
QY 1829 ADALERFGENRFVSPLSWDLVGRNLFAMAVEGVFLITVLIOYREFIRRPVNAKLSP 1888

```

```

Db 1854 TDVVAQGEESANPPOWDLGKNLVMAIEGVVYFLTLTLLIHHFLLRMIAEPAREP 1913
QY 1889 NDEDEDVAREPORILIDGCGNDILEIKELTKIYRRKKRPADVIRCVIIPGECFGLGVN 1948
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1914 FDEDDVAAERQKRVMSGKNKTDILKLNELTKYVSGSSSPAYDLVCYVRGECFGLGVN 1973
QY 1949 GAGKSTTEKMLTGTITVYRGDAFLINKSILSNIEHVONNNGYCPORDATITELLGRENH 2008
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1974 GAGTTFEKMLTGTITVYSGDATVAGKSILTSISDVQNNNGYCPORDADITELLGRENH 2033
QY 2009 FFALLRGVPEKEVKGEMAIRKLGKVKYGVKAVNNGSGNKKRSLTAMALIGGPVVL 2068
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2034 LYARIRGVPSKEIKVANNWGIOSIGLSLYADRLAGTYSNGNKKRSLTALITGCPULLL 2093
QY 2069 DEPTTGMDPKARFELNMCALSVNEKRSVYLTSNMECEALCTRMAIYNVGRFRLGSV 2128
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2094 DEPTTGMDPKARFELNMCALSVNEKRSVYLTSNMECEALCTRMAIYNVGRFRLGSV 2153
QY 2129 QHLKRRGDTTIVYRAGSN---POLKRYODFFGLAFGVSULKEKRNMLDOLPSSL 2184
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2154 QHLKRRGDTTIVYRAGSN---POLKRYODFFGLAFGVSULKEKRNMLDOLPSSL 2211
QY 2185 SSLARITSILSOSKKRLHIEDYSQTTLDQVFNFAKDDSDDLKDLISLH 2236
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2212 SSLARITOLISHKDSLIBEYSVTQTTLDQVFNFAKDDSDDLKDLISLH 2260

```

RESULT 5

```

US-09-995-542-12
Sequence 12, Application US/09995542
Patent No. US20020127647A1

```

```

GENERAL INFORMATION: John
APPLICANT: Shuttle, John
TITLE OF INVENTION: ATP-Binding Cassette Transporter-Like Molecules and
FILE REFERENCE: 00-658-A
CURRENT FILING DATE: 2001-11-28
PRIOR APPLICATION NUMBER: 60/253,520
NUMBER OF SEQ ID NOS: 24
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 12
LENGTH: 2273
TYPE: PRT
ORGANISM: Homo sapiens
US-09-995-542-12

```

```

Query Match 49.6%; Score 5854.5; DB 10; Length 2273;
Best Local Similarity 49.9%; Pred. NO. 0;
Matches 1163; Conservative 356; Mismatches 637; Indels 171; Gaps 27;

QY 6 QURLIMKNLTPFRROTQCLLEAVAMPLEFLILISVRLSYRSPYEOHECFPNKAMP5AG 65
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 6 QURLIMKNLTPFRROTQCLLEAVAMPLEFLILISVRLSYRSPYEOHECFPNKAMP5AG 65
QY 66 TLPPVQGITCANNDPCRYPPRPGAPGVGKFNKSIYARLEFSARLLIYQKDTSKDM 125
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 66 TLPPVQGITCANNDPCRYPPRPGAPGVGKFNKSIYARLEFSARLLIYQKDTSKDM 125
QY 126 RKVLTLOI-----KSSSNLKLQDPLVNETFSGFLYNNLSLPKSTV---DKM 172
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 126 RKVLTLOI-----KSSSNLKLQDPLVNETFSGFLYNNLSLPKSTV---DKM 172
QY 173 LRADVLTKRVLQGYOHLITSL-NGSKSEMIOLGQEVSE---LCGLPREKLAAB 226
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 173 LRADVLTKRVLQGYOHLITSL-NGSKSEMIOLGQEVSE---LCGLPREKLAAB 226
QY 186 VRPEGFANGV---PDLATKDIAECSEALIERILFSGRGKATYRVALCSLSQGTQWIE 241
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 186 VRPEGFANGV---PDLATKDIAECSEALIERILFSGRGKATYRVALCSLSQGTQWIE 241
QY 227 RVLRSNMILKPLIKTLNLSSTPFSKELAEATKTLHLASLGLAOLFSMSGW---SDM 281
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 227 RVLRSNMILKPLIKTLNLSSTPFSKELAEATKTLHLASLGLAOLFSMSGW---SDM 281

```

QY 282 -----ROEVEFLTNV-----NSSSSSTQIOAVSRIVCGHPEGGLKIKSLNWE 326
Db 282 PRIOEIHRPSMODLWVRPLMONGCPETFTKIMGILBDLIGCYEGSGSVLSTNWE 341
QY 327 DNNYKALFGNGEAEAEFFYNSTPTPCNDLMKNLESSPLSRITWKALKPLLVGILYT 386
Db 342 DNNYKALFGIDSTRKDPISYDRRTSPFNALQSLSPKIMRAKPLMGLKILYT 401
QY 387 PDTPAROVAAEYKTFEOELAVFHDEGMMWELSPKIMFMENSOEMDLYRMILDSROMD 446
Db 402 PDSPARRILKANSTFEELHNRKLYKAMEYGOIYWFNDNSTOMNTRITLGLPYVK 461
QY 447 HEWEQOLGLDWTADIVAFKAKHPEDVOSNGSVYTWREAFENETNOAIRTISREMECVN 506
Db 462 DEFNROLEGEGITAEAILNFIKYGPRESOADDMANDFMDINIDRTLRVNOVLECLV 521
QY 507 LNKLEELAEVWLINKSMELDERKFWAGIVFTGTPGSELPHHYKXIRMDIDNVERT 566
Db 522 LDFEESTYNDETQTORALSLLENMFMAVPEPDMYPMWTSLSLPHHYKXIRMDIDVERT 581
QY 567 NKIKDGYWDPGRPADPEFEDMRYVWGFAFYLODYVEAOITRVLGTGKTKGYVMQMPYC 626
Db 582 NKIKDGYWDPGRPADPEFEDMRYVWGFAFYLODYVEAOITRVLGTGKTKGYVMQMPYC 641
QY 627 YVDDIFLRWMSRMLPMTLWAMYVAITIKGIVYEKARLEKIMGLDNLISLWFW 686
Db 642 FVDDSMFILNRCEPIFVWLMYISVMTKSIKLEKELKELKNGVNAVIMCTWF 701
QY 687 ISLLPLVASAGILVYITLGNILPYSDBSVYFVLSTFAVNTILQCLISLISFRANLA 746
Db 702 LDFEESTYNDETQTORALSLLENMFMAVPEPDMYPMWTSLSLPHHYKXIRMDIDVERT 761
QY 747 AACGGIYFTLPLVYLAVYAMODYVGTFLKIPASLSPAFGCEGYFALPFEQIGYVOW 806
Db 762 AACGGIYFTLPLVYLAVYAMODYVGTFLKIPASLSPAFGCEGYFALPFEQIGYVOW 821
QY 807 DNLFESEVEDGFNLTSTISMLPFTFLYGVMTWYIEAVFGQYGI PREWYPCRTKSYWF 866
Db 822 SNIGNSPTGDEFSLLSQMMLLDAAYVGLAWYLDQVFGDGYTPPLWYFLDESYWL 881
QY 867 G-----EESDEKSHPGSNOKHSEICMEEPETHLKLGSITONLVK 907
Db 882 SEGGCTREERALEKTEPTEETDEBP-----EGIHDSFEERHPGVGVANLVKI 937
QY 908 YRDKMVAVDGLALNFEGQITSFLGNAGAKTMTSILTGLEPPTSGTAYILGDIRSE 967
Db 938 FEPGGRPAVDRLNTEFYENQITAFLGHNAGAKTMTSILTGLEPPTSGTAYILGDIRSE 997
QY 968 MFTIRONGVCPONHVLFTMLTVEEHIMYARLKLSEKHVAEOMALDYGLEPSSKLK 1027
Db 998 LNAVROSLGMCPOHNLFTLHNLVAEHMLFYAOLKGRSOEAOLEMEAMLEDTGL-HKRN 1056
QY 1028 SKTSQSLGCMQRKLYALAFVGSKVILDEPTAGVDPSPRGIMELLILKYGKGTITLS 1087
Db 1057 BEADOLSGMQRKLYALAFVGSKVILDEPTAGVDPSPRGIMELLILKYGKGTITLS 1116
QY 1088 THHMDADVGLRIALISGKLCYSSSIFLKNQJGTGYTLVVKDYESSLSRCRNS 1147
Db 1117 THHMDADVGLRIALISGKLCYSSSIFLKNQJGTGYTLVVKDYESSLSRCRNS 1175
QY 1148 TVSYLKREDSVSSSDGLGSDHSDPTLIDVSAISNLIRKHVSEARLYVEDIGHETLYV 1207
Db 1176 TSCSSKGS---TTCRAVDDLTPEOYLDGVDVNEMLDVHNEVAKLVEIGELFEL 1232
QY 1208 LPEYAKBEGAVFELFHEIDRLSDLGISYSIGISFTLLEIFLVAEESGVDAETSDGTLR 1267
Db 1233 LFNKNFKHRAVSLRLELETLADIGLSSFGISDTRPLEIFLKVIEDSDSGFLRFG--- 1289
QY 1268 ARNRNRAREDKSCLRPFTEDAADPNDSIT-----DPESREIDLLSGM 1311
Db 1290 AOKKRENNVRRHPCIGP-REKAGQTPDSSNWCSPAPAHNPGGPPPEPCGQLNLTGT 1348

QY 1312 DGKSYGVKGMKLTQOQFVALLMKRLIARRSRKGFPAQIVLPVAVCIALVSTLIVPE 1371
Db 1349 -----QVLQHOQALLVRFQHTTISKHDPLAQIYLPRTVFLAMLSVTLPL 1397
QY 1372 GKPSLELQPMWYNEQYTFVSNDAPEDTGTLELLNATLKDPFGFTROMEGNPIDPFCQA 1431
Db 1398 GEYALTLHPWLYGOQYTFVSNDAPEDTGTLELLNATLKDPFGFTROMEGNPIDPFCQA 1456
QY 1432 GEEETAPVYQITMDLFGONNMTMOPNSPACOSSDKIKMLPVCPPGAGLPPOPKQ 1491
Db 1457 NSTPKTPTSVSPNTQTLFQOKMTQVNPSPCSTREKLTMLPECEGCGAGLPPOPKQ 1516
QY 1492 NTADILQDLTGRNISDYLVKTYVOIISKLNKIMVNEFRGGSGLVSTNOALPPOQEV 1551
Db 1517 RSTELQDLTGRNISDYLVKTYVOIISKLNKIMVNEFRGGSGLVSTNOALPPOQEV 1572
QY 1552 NDAIKOMKHLKLKSSADRFNLISGR-----FMTGLDTRNNKV 1592
Db 1573 GEALV-----GFLSDLGRIMVSGGPITREASKETPDPLKHLTEDNITV 1617
QY 1593 WPNKGMHAISSPLVNTNALLRANLKGEMPSHYGTAPRHNPLNLTQOLSEVALMTTS 1652
Db 1618 WPNKGMHAISSPLVNTNALLRANLKGEMPSHYGTAPRHNPLNLTQOLSEVALMTTS 1677
QY 1653 VDVAVSICVIFAMSEFVAPFVFLIOERVSQAKHLOFISGVKPYVWLSNFMVDMCVYV 1712
Db 1678 VDVAVSICVIFAMSEFVAPFVFLIOERVSQAKHLOFISGVKPYVWLSNFMVDMCVYV 1737
QY 1713 PATVYIIFTCFOOKSYVSTNPLVALLLIGWSTPLMYPASFEVKIPSTAYVLTTS 1772
Db 1738 SAGLVAVGIFGOKKAYTSPENLPAVALLLYGMAYIPMYPASFLDVPSTAYVLTTS 1797
QY 1773 VNLFGINGSVATFVLEFETDNK-LNNINDLKSVLIFPHECGRGLDMVNOQAMADA 1831
Db 1798 ANLEFGINSATITLLEFEDNKRLLFRNVALKRLLVFPHFCGRGLDLSALQATDV 1857
QY 1832 LERFGENRFPVSLMDLVGRNLFAMAEGVVEFLITVLIQYRPIRPPVNAKLSPLNDE 1891
Db 1858 YARFGEESHPAFHMDLIGKMLFAMVGEVYVFLITLLOVHNFLLSQWIMAPKEPIYDE 1917
QY 1892 DEDVAREKORILDDGGONDILEIKELTKYIRKRAKRAVDRIKCIIPGECFGLGVNAG 1951
Db 1918 DDDVAEERORILITGKNKTDILRLHELTKIYLGTSPPAVDRICVGVNPGECFGLGVNAG 1977
QY 1952 KSTFEMKLTGDTVTRGDAFLNKSILSNIEHYHONNGYPOPOALITELTGREHEEFA 2011
Db 1978 KTTFFKMLTGDTVTRGDAFLNKSILSNIEHYHONNGYPOPOALITELTGREHEEFA 2037
QY 2012 LLRGVPEKGVGEMAIKRLGLVYGEKYAGNYSGNKRKRLSTAMALLGGPPVFLDEP 2071
Db 2038 LLRGVPEKGVGEMAIKRLGLVYGEKYAGNYSGNKRKRLSTAMALLGGPPVFLDEP 2097
QY 2072 TTGMDPAKRFPLNCAISYKSEGSVYLTSHSMECCALCTRLAIVNGKFRCLGSQHL 2131
Db 2098 TTGMDPAKRFPLNCAISYKSEGSVYLTSHSMECCALCTRLAIVNGKFRCLGSQHL 2157
QY 2132 KNRFGDGYTVIVRITAGSN---PDLPVODPFGFLAFEGSLKKEKRNHMOYQAPSSLSL 2187
Db 2158 KNRFGDGYTVIVRITAGSN---PDLPVODPFGFLAFEGSLKKEKRNHMOYQAPSSLSL 2215
QY 2188 ARITSISQSKRKLHIDTYSQTTLDQVYVNRPAKDOSSDDHLKDLJLH 2236
Db 2216 ARIFOLLSHKDSLILBEYSVTQTTLDQVYVNRPAKDOSSDDHLKDLJLH 2261

RESULT 6
US-09-995-542-5
Sequence 5, Application US/09995542
Patent No. US20020127647A1
GENERAL INFORMATION:
APPLICANT: Shutter, John
APPLICANT: Uljas, Jaani
TITLE OR INVENTION: ATP-Binding Cassette Transporter-Like Molecules and

Oy	829	LEDFELGVMWTVLEAFPCOYQJPRPWRPCWSEFEESDEKSHGNOKRMSEST	888
Oy	839	LEDFELGVMWTVLEAFPCOYQJPRPWRPCWSEFEESDEKSHGNOKRMSEST	888
Db	735	LLDAVITGLAMTIEANVCPOYQJPRPWRPCWSEFEESDEKSHGNOKRMSEST	793
Oy	889	EEEPHTLKLGVSTONLVKVVYRDMKVAVDGLALNFEGOITSPLCHNGAKTTTMSLTG	948
Db	794	EEEPHTLKLGVSTONLVKVVYRDMKVAVDGLALNFEGOITSPLCHNGAKTTTMSLTG	853
Oy	949	LEPPTSGMAYLKDRJSEMSSTRONLGYCQPNVTFDMLVYEBHTWYARLKGISEKHV	1008
Db	854	LEPPTSGMAYLKDRJSEMSSTRONLGYCQPNVTFDMLVYEBHTWYARLKGISEKHV	913
Oy	1009	KAEENOMALDVGLEPSSSKLKSCTQLESGMQRKISYLAFAVGSKVILDEPAGDPSR	1068
Db	914	GPERERILRDVGL-TLKRDTQHTLHSGMQRKISYLAFAVGSKVILDEPAGDPSR	972
Oy	1069	RGIMELLKTRQGTITLSTHMDADVLGDRJAITSHGLCCVSSPLFKNQJSTGYLL	1128
Db	973	RGIMELLKTRQGTITLSTHMDADVLGDRJAITSHGLCCVSSPLFKNQJSTGYLL	1032
Oy	1129	TLVKKDYESSLSSCRNSSVSYLTKEDSVSSQSSDAGJGSDHESDT-----	1175
Db	1033	TLVKKDYESSLSSCRNSSVSYLTKEDSVSSQSSDAGJGSDHESDT-----	1081
Oy	1176	-----LTIDVSAISNLIKRHHSEARLVEDIGHETLVYLPKAEKEGAFVELFHEIDR	1228
Db	1082	PAPGAVPTPSTAIETELVYOHVPAQVALVEDLPHETLVYLAAGLDSFAMVPELDQ	1141
Oy	1229	LSDJGSSYGSSETELEITFLKKAESSGVDAETSDGTLPARRNRAFCK--QSTLRPT	1266
Db	1142	LELGLGLGYGSDTNELEITFLKKAESSGVDAETSDGTLPARRNRAFCK--QSTLRPT	1185
Oy	1287	EDDADADENDS-----DIPESRETDLSGMGKGSYVQKMWKLTQOQFVALLMKR	1336
Db	1186	POPPTGEASYLENGELAKLYLDRQAQ-----GL-APNAAVQGVMTLTCOOLRALLMKR	1239
Oy	1337	LLIARRSRKGFPAQIVLPANVCIALVFSILVPPGCKTIPSELEQPMWNTBOYTFYSNDAP	1366
Db	1240	FLIARRSRKGFPAQIVLPANVCIALVFSILVPPGCKTIPSELEQPMWNTBOYTFYSNDAP	1299
Oy	1397	EDTGTLELNAITFDPOGFTGRCEGNAPIDPTPOAGEEEMWTAP--VPQTIMDLFONGWTT	1455
Db	1300	GDPNRMKLLEALLGEOLOPMSQMDADAGSECTHSLACTIYFVEYPRDVAIILASGWT	1359
Oy	1456	MONPSPACOSSDKIKKMLVCPPGAGGLEPPQOKONTADILQDLTGRNISDYLVKTYVQ	1515
Db	1360	PESPSPACOSSDKIKKMLVCPPGAGGLEPPQOKONTADILQDLTGRNISDYLVKTYVQ	1419
Oy	1516	IIAASLKNKIMVNEFRRGGSLSGVNTQALPSEOEVNDAIKOMKHKLILANDSSADRPLN	1575
Db	1420	LVRRGLTKTKMWDVRRGGSLSG-GRDPDLPGHEVVRTLAEIRALLSPQONADIRLNL	1478
Oy	1576	SLGRFMTGLDTRNNVAKWFMNKGMAHISSEFLINNAIILRAMLOKGENBSHGTAFANHP	1635
Db	1479	NLTQMALGLDARNSIKIMFNKGMHMAVAFVNRANNGLLHALLPBGVPRHAHSITTYLHP	1538
Oy	1636	LNLTKOOLEVALMTTSDVLVSIQVIFAMSPVPAFSEVVELTQERNVSKAKHLQFTSGVKP	1695
Db	1539	LNLTKOOLEVALMTTSDVLVSIQVIFAMSPVPAFSEVVELTQERNVSKAKHLQFTSGVKP	1598
Oy	1696	VIVYMSFNWDMCMVYVPAVLVILIFELPCOQSYVSTNLPVIALLLLLYGSITPLMWP	1755
Db	1599	TLVYAGNFMWDMCMVYVPAVLVILIFELPCOQSYVSTNLPVIALLLLLYGSITPLMWP	1658
Oy	1756	ASPEVKITPSTAYVVLVSNLFEJGINGSVATVELEFTDKLNNINDILKSVFLFPHCL	1815
Db	1659	ASPEFVSPTAYVVLVSNLFEJGINGSVATVELEFTDKLNNINDILKSVFLFPHCL	1718
Oy	1816	GRGLIDWVKNOAMADLERGENRFPVSLSPMLJQGNLFLAMAVEGVFFLTVLQYRFF	1875
Db	1719	GRGLIDWVKNOAMADLERGENRFPVSLSPMLJQGNLFLAMAVEGVFFLTVLQYRFF	1778

Qy	1876	IRPPVNAKLSPLUNDEEDVREPRROGILDDGGONDIETKELTKYIKRRKRPADVRIOW	1935
Qy	1779	LLPOSKRLLPLRGEEDEDVQREKRYTKGATOGDVLVRLDTKKYKQOQRPADVRLCLG	1838
Qy	1936	IPPECGLLGVNGAGKSSSTFKMLTDDTVTRGDAFLUNKNSILSNIEHVHONMGYPOFD	1995
Db	1839	IPPECGLLGVNGAGKSTFRVNTDITLPSSEAVLGHANVAOERSAHHSMGTCPSD	1898
Qy	1996	AIETELLTGREHVEFFALLRGRPEKVEKVGEMAIKKLGLVKYKGYAGNYSQGNKRKLST	2055
Db	1899	AIFOLLTGREHLEFALRLRGVPEAOVAQJALSOVLBGLPSYADRPAGNYSQGNKRKLAT	1958
Qy	2056	AMALIGRPVVELDEPTTGMDPKAREFLMNCALSVYKERSVYLTSHSNEBCALCTPMA	2115
Db	1959	ALALYGPRAVVEFLDEPTTGMDPSAREFLMNSLTSVYREGRSVYLTSHSNEBCALCTPMA	2018
Qy	2116	IMVNGRRCIGSVOHILKRNRCGDYTLVYVLAGSNPDLKRVQDFOLAPRGSVILKXHNMM	2175
Db	2019	IMVNGRRCIGSSOHLKGRFGAGHTLTLKRVPPDQF-PALAFIRITPPGALREVHGR	2076
Qy	2176	LOYOL-FSSLSLARIPSLISOSKRRHILIEDYSVOSTLTDQYFVNMADQSDDD	2228
Db	2077	LRFOLPQGGKCTLTVRYRELALAOGRAGHVEDSVSOTLLEEVFLTFSSKQDQGE	2130

[illegible]

SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 3
 ; LENGTH: 2121
 ; TYPE: prt
 ; ORGANISM: Mus musculus
 US-09-595-542-3

Query Match 47.3%; Score 5581; DB 10; Length 2121;
 Best Local Similarity 50.2%; Pred. No. 0;
 Matches 1118; Conservative 350; Mismatches 570; Indels 190; Gaps 32.

47 PYEHCHEPFRKAMPAGTLPPWVQIICNANPCFRYPTEAPGVGNFKSIVARLE 106
 1 PLEHHCHEPFRKAMPAGTLPPWVQIICNANPCFRYPTEAPGVGNFKSIVARLE 60
 107 SDAKRLIYSQKTSKMDKRVLTLOQIKKSSNLKLODFLVNDETSGFLYHNLILPK 166
 61 ADRTYVL---GGHSIQDMDLAGKLLPY----- 85
 167 STYDKMLRADVILHKVFLQGYQLTSLGNSKSEEMIQDOEVSGLCPKREKLAAE 226
 86 -----LVAVGGAAPQF---SDQPTSQ---GSYTKLL---E 112
 227 RVL-RSNMDILKPIELTNTSTPFPKSELEAKTLLHSLGTLAOLFFSRMSNDKROEV 285
 113 KILORASLD---PYL-----GQADSMKRFSDAIRDLAQELLTLPISLMELR--- 155
 286 MFLTNSSSSSTOIQVANSRYC---GHPREGGLKIKSLNWNEDNNYKALFGNGTEED 342
 156 ALLRPRPSAGSLLEL---VSEALCSTKPSPEGL--SLNWEAQLWNEFMG---PREV 205
 343 AETFYNSTPYCNDMLKMLESSPLRIIMKALPLVGLKIITPTPATROYMAEVNKT 402
 206 APRLPDNLSPACSEFYGLTDHPVSRILMRKRLKILFAPRTNTRKMAVNGT 265
 403 FOELAVFHLEGMWELSKITWTFMENSOEMLVRLMLDSRDNHFWEOGLDMLTAOD 462
 266 FEELALRLDLHLEMGVGLQIFENFMDSTVAVMLQRLLYVGTCG--ROOTPRAKKLEA 323
 463 IYVFLAKHPEDVOSSNGSVYTWREAFNETNOAIRTISRPECVNLKLEPIATEVYVLLK 522
 324 IKQFL-----DPSRGG-YSMREAHADMGRLAGITGOMMECSLKLLEAVSEELVSR 375
 523 SMELLERKFWAGIVFTG---ITPQSID---LPHVHKYKIMRIDNVERTNKIKDGYW 574
 376 ALLELGERRLMAGIVFLSPREHPLDSPSLSPALSPGHLRPIKIMDDVYTRNKIKDGM 435
 575 DPGPRADPEEDMKRYWGGGAYLQDVVEQALITRYLTGTEKTYVMOOMPTPCYVDIPLR 634
 436 DPGSADPEFMDLRYWGGGAYLQDVVEQALITRYLTGTEKTYVMOOMPTPCYVDIPLR 495
 635 VMSRSMPEFMTLAMIYSVAIIKIYIYKEARLKETMRIGLNSILMFSWIFSLPL 694
 496 VLSRSLPLFLTLAMISVALTVKAVAREKTRIRETRAMGLSRAYLWGLFSLCGLPFL 555
 695 VSAGLVLVILKLNLLPYSDPSVVEFLSVFAVVTLLQCLFISTLFSRANLAAACGIIY 754
 556 VSAALVLVILKLNLLPYSDPSVVEFLSVFAVVTLLQCLFISTLFSRANLAAACGIIY 615
 755 FTLVLPVLCVAMODYVGFLLKIFASLISPVAFGCEYALFEEOGIGVQMNLFESPV 814
 616 FALYLPVLCVAMODYVGFLLKIFASLISPVAFGCEYALFEEOGIGVQMNLFESPV 675
 815 EEDFNLTTTISMLFDTFYLGVTWVTEAVPFGQYGIIPRPMYFPCPKSWPFEESEDEK 874
 676 -EDVFSIAQVSAFLLDLAYITGLALVLEAVCGQYIIPRPMYFPCPKSWPFEESEDEK 733
 875 HPGSNORMSEICMEEPETHLKLGVSTIQNLVKKYRDMKAVADGLALNFYEGQITSLGH 934
 734 SYLAPAPQDKRVLVEEPLGLVPGVSTIRGLKHKHRCGPCPALGCLNDFEYCHITAFGLH 793
 935 NGAKTTMTSLGLRPTSGTAYILKODIRSEKSTIRQLGLVCPQHNVLFDMLTVEEH 994

794 NGAKTTMTSLGLRPTSGTAYILKODIRSEKSTIRQLGLVCPQHNVLFDMLTVEEH 853
 995 WFYARLKLSEKHVKAEMOMALDVGLPSSKLSKTSQSLSGMORKLSVALAFVAGSKV 1054
 854 WFYGRKLVSAAMGPREERELIRDYGL-TLKRDTQTHLSGGMORKLSVALAFVAGSKV 912
 1055 ILDEPTAGVDPPYSRKGIWELLKLYKQGTITLSTHNDDEADVLGDIRAIIISHKRLCVGS 1114
 913 IMDEPTAGVDPPYSRKGIWELLKLYKQGTITLSTHNDDEADVLGDIRAIIISHKRLCVGS 972
 1115 SLELKNOLGTGYTLTKKKDVESLSCSRNSSSTVSTLTKEDSVSSQSDAGLSHED 1174
 973 PLFLRHLGCGYVTLVK-----SSQSLVTADKADSDDPRREKSDNGRTST 1021
 1175 T-----LTIDVASISLIRKHYSEARLVEDIGHELTYVLPYPAK 1214
 1022 TAFTRGTSDDKNQAPAGAVTTPSTARILELVQOHVPGADVLDLHLELLVPLTAGAL 1081
 1215 EGAPELFEHIDRLSDGISISYGETTELEIFLKAEEGVADETSDGLPARNRRA 1274
 1082 DGSFAMVFOELDQELLEGLTGYSIDTNLEIFLKYVEDA-----HRE 1125
 1275 FGDK--QSCLRPTEDDAADPDS-----DIPSERETDLSGMDKSGYQKGV 1322
 1126 GDSRQPLHLRTPCTPQPTGPEASYLENGELAKLVLDPOAPO-----GL-APNAQVQGV 1179
 1323 KLTOOQFVALLMLKRLIARRSRKGFPAQIVLPAVEVCALVFSLIVPPGKPSLELOPW 1382
 1180 TLTQOQFVALLMLKRLIARRSRKGFPAQIVLPAVEVCALVFSLIVPPGKPSLELOPW 1239
 1383 MYNQYTFVNSDAREDTGTLLELNLATKDPGFGTRCEGNDIPDTPOAGEEWTAP-V 1441
 1240 MYGPQVSEFSEEDAPDPRMKLLEALIGEAGIOPESMDKARSECTHSLACYFTVEV 1299
 1442 PQTMDLFQNGWMTQNSPACQSSDKIKKMLRYCPRGAGLIRPPQKQNTALIDDT 1501
 1300 PRDVASILASGWMPESSPACQSSDKIKKMLRYCPRGAGLIRPPQKQNTALIDDT 1359
 1502 GNISDYLVKTVQVQIITAKSLNKIWNNEFRYGFSLGVSNTQALPPSOEVDAIKQKKH 1561
 1360 GNVSDPLVTKYPSLVRRGLTKKKWDEVRYGSPSLG--GRPDLPTGHEVRYTLAETRAL 1418
 1419 LSPQGNALDRILNLTQWALGLDARNSLKIWFNNKGHWAAVYVNRANNLHALLPSG 1478
 1622 EMPSHYGITAHRNLTNLTQWALGLDARNSLKIWFNNKGHWAAVYVNRANNLHALLPSG 1681
 1479 PVRHAHSTTLNHLNLTREKQLEATVTLASSVDVLSICVVPFAMSFVPASTLVLEBR 1538
 1682 SKAKHLOFISGVKPIYVLSNFEVMDCMYVVPATLVIIIFCFOOKSVSSTNPLVAL 1741
 1539 TRAKHLOVLSGVKPIYVLSNFEVMDCMYVVPATLVIIIFCFOOKSVSSTNPLVAL 1598
 1742 LLLYQWSTPLMYPASVFKIPSTAYVVLVTSVNLFTGINGSVATFVLELPTDNKLNIND 1801
 1599 LLLYQWSTPLMYPASVFKIPSTAYVVLVTSVNLFTGINGSVATFVLELPTDNKLNIND 1658
 1802 LLSVFLFPHHCLRGILIDWYKNOAMADALERGESENFVPSLSDVIGRMLPMAVAGV 1861
 1659 LLSVFLFPHHCLRGILIDWYKNOAMADALERGESENFVPSLSDVIGRMLPMAVAGV 1718
 1862 VFLITVLYQVFFRPRPVNAKLSPLNDEDEDVYRERORILDDGGGNDILEIKELTY 1921
 1719 LFLITVLYQVFFRPRPVNAKLSPLNDEDEDVYRERORILDDGGGNDILEIKELTY 1778
 1922 RRRKRPADRICVGIIPGECFGLGVNGAGSSTFKMLTGDTTYTGRGAFANKSITLSNI 1961
 1779 RGRNPADVDRICVGIIPGECFGLGVNGAGSSTFKMLTGDTTYTGRGAFANKSITLSNI 1838
 1982 HEVHOMGVCPQGFDAITELTGRHVEFPALLRGVPEKGVCGEMALYRKGLVYGGKY 2041
 1839 SAHRSMGYCPQGFDAITELTGRHVEFPALLRGVPEKGVCGEMALYRKGLVYGGKY 1898

Qy	1170	DHSDFTLIDVS-	-ALSNIIRKHSV	ARLVEIDIGHELT	YVIVPAEKAGPAVLFHEID	1227		
Db	1239	SPGRAPLSSCSELO	SOFTRKHVASCLLV	SDTSTELSTILP	SEAKKAGAEERLFÖHLER	1298		
Qy	1228	RLSDUGISYSGETT	LELEFLKVADE----	SGVDA-	ETSDGTLP-----	1267		
Db	1299	SLDLHLSSFGMLD	TTLLEBFLKVSDEB	DSLENSEADYKES	RKVDLPGAEPSAGCEGHAG	1358		
Qy	1268	-----	ARRNRRRA-	FGDKOSCLRP	TTEDDADPRDSDIDPESR	1303		
Db	1359	NLARCSELTO	SQASLOSASSVGS	SARDEBAGTVDY	GYDRPLF-DNPDDP--	NVSLQEV 1415		
Qy	1304	ETDLSDGMDGCS	QVOYGMWLTQOQ	EVALMLKRLILARR	SKRGFAQOIVLPAVE	CIALV 1363		
Db	1416	EAEALSKY-	QOGSKRLDGG	LKXKROFHGILLV	RPHCARNSKALP	ESQILLPAPFVCAMT 1474		
Qy	1364	FSLIVPEPGKYP	PSLEILOPMYIN	EOYT-----	FVSDAPF-----	DTGTLELLNAL 1408		
Db	1475	VALSPPELGDPL	PVLPVLPSSQYH-	NYTOPRGNEFLPY	ANERREYRULR	SPDASPPOOLVSTF 1533		
Qy	1409	TKDQGFGRCH-	-----	ECAPFI-----	-----	1424		
Db	1534	RLPBGVATCYL	KSPANGSLGPTL	NSGSESLAARF	DSMCLESPFOGUEPL	NFVPP 1593		
Qy	1425	-----	PD-----	TPCQAGEE	BWTAP-VQITMDLF	ONGNMTMONPSPAC 1463		
Db	1594	PSPAPSDSPAS	PDEDLOAMVWS	LPPTAGEBMT	SAPSLERLREBPVR-----	-C 1641		
Qy	1464	QCSDDKIKKML	PVCPBPAGGLP	PPROKONMADLO	TGRNLSIDLV	TYQOIIAKSLKN 1523		
Db	1642	TCSAQGTGFS-	-CPSSVG-	HPPOKRVVTG	DLTDITGHANSEY	LEFTSDRF----- 1690		
Qy	1524	KIWNEEFRYGG	FSLGVSNTALP	PPSOEVDNDAIK	OMKHKILAKSDS	ADRFNLSLGRFMTG 1583		
Db	1691	----	RLHRYGATITGC-	NVLKSI	PASPGFRAPMVRK-----	1721		
Qy	1564	LDTRNNK	YVNNKNGMAIS	FLNVALNNA	ILRANLOKGE-NPS	SHGITAFAHPHPLNTMQO 1642		
Db	1722	IAVRRAQV	EFYNNKGYISM	PTLNSLNNA	ILRANLPRKSGN	AAAGITVTNHPNPKMTSAS 1781		
Qy	1643	LSEVAMLT	SDVLSVJCVIF	FAMSEVPAS	FVFLIOERYSKAKH	LOFTISGVPRVYWLNS 1702		
Db	1782	LS-IDYLOG	CDVYIAFLI	YAMSFPVPS	VYVFLVAEKSTK	KHLOFAGSCNPIIYMLAN 1840		
Qy	1703	FVDMC	NVBPATLVITIT	FCFOOKSV	SVSTNLPLVAL	LLLLLXMSITPLMPASVEKI 1762		
Db	1841	YVMDL	NLVNLPATCC	VIILIEVFDLP	ATPSTPNPA	LSLELLGWSITPIMPASFWEV 1900		
Qy	1763	PSTAVVLT	SVNLFIFIGNS	AFPFVLELTP-DNK	LNNINDILKS	VELLJPHHCLOGRGLID 1821		
Db	1901	PSASVAV	FLITNLFTGIT	VAIVAFLLLO	LFPHDLDLVVNS	YLSKCELLPBNLHGHME 1960		
Qy	1822	MKNOMAD	ALEREGE-NRE	YSPLSMDV	GRNLFFAMA	VGVSFFELTYVLIQYRPIRRP 1880		
Db	1961	MAYNEY	INEYAKIG	QEDKMKSP	PFEMDI	VRGLVAMA	VGVOFLITIMCOYNFLRRPQ 2020	
Qy	1881	VNAKSL	PUNDEDEV	RRRQRJID	GGGONLIEIK	ELTYIRRK---KRP	ADRLCVGP 1937	
Db	2021	MVSTR	PEVD-DVIV	ASROVFL	RGADNDNMV	ITENLITVY	YSKRIGRLAVDLRCLCVR 2079	
Qy	1938	PECEFG	LGVNGACK	STFKMLT	DDTVTR	RGDAFLK	NSILSHIEHVONMGYCPQFOAI 1997	
Db	2080	PECEG	LGVNGACK	STFKMLT	DEBTTG	GEAFVNG	HSVSKELLDOYOSLOTCPOCAL 2139	
Qy	1998	TELLGRE	VEFPALLRG	VEPEYK	GVEMAKRLGL	VKYEGER	YAGNVSNGKRLSTAM 2057	
Db	2140	PELTA	REHLOLYR	LROISK	WDARVYK	AKALEKLELT	KYADKPA	TYSGGAKRLSTAI 2199
Qy	2058	ALIG	BPVAVLE	DEPTGDDP	PARFELN	CAVYK	GBRSVVL	SHSMECCALCYRMAIM 2117
Db	2200	ALIT	GPARTFL	DEPTGDM	PARFELN	LILDLIK	GRSVYLT	SHSMECCALCYRLAIM 2255

[illegible]

```

QY 493 -----QAIRIS-----RMECVNLKLEPLATEWMLINKSM-----LDERKFNAG 535
Db 548 LPSGMALLOQDLTDIDNACGMIOFMKSVSDIFKGFPEDESIYNYTLNAYQONVAFAS 607
QY 536 IYFTGILPGSIELPHYKXKIRMDINVERTNKIKGYNDDPGRAPFEDMRVWGCFAY 595
Db 608 VITOTRKDS--LPHVHKIKRONSSFTKENTIRAYWPKPNTG--GREFFLYGFWM 662
QY 596 LODVDOAILRVLGTE-KKTGVYMOOMPYPCYVDIPLRVMSRMPILMTLAMIYSAV 654
Db 663 IODMERAITDITFVGHVDPGSGYQMPYPCYTRBDPLFJEHMMPLONVSWYSAVM 722
QY 655 IIKGIYERKARLEKTRIMGLDNSTLWFSLSLPLLVASAGLVLLKGNLPSD 714
Db 723 TIOHIAEERHLEKVMKTMGNNAVHVAWFTGVOLSISVTLATLILKGVYLHSH 782
QY 715 PSVAVFELSAFVAVTILLOGLISTLPSRANLAAAGGIIYFTLYLPLYVC-----VAMOD 769
Db 783 VYIMFLAVYAVATIMFGLVSVLSKAKLASAGGIIYFTLSYPMYVAIREVAHDK 842
QY 770 YVFTLKIFASLLSPVAFGECEYFALPEBOIGVOMDMLFESPYEDGFNLTSISM 829
Db 843 ITAFE-KCIASLMSTAFGIGSKYFALYEVAGVGIOMHFFSOSPYEGODFNLLAVTLM 901
QY 830 FDFELGVMTWYEAFFPOQYGIIPRWYPTCKSYFGE--ESDEKSHPGSNOKRMS-- 884
Db 902 VDAVAVGILWYIEAVHPQMYGLPRPWYPTLOKSYWLSGRTAEAWMSWPMARPTLSYM 961
QY 885 ---EIC-----MEEPHILKGVSIONLVKVRDGMKVAVDLALNFEQGT 929
Db 962 EEDQACAMESRREETRGMEETRHPLVVCYDKLTKYKDKKIALINKSLINLEYNOV 1021
QY 930 SEFGHAGAKTTTMSLJGLFPPTSGTAYILCKDIRSEKSTIRQMLGYCPOHNVLEMLT 989
Db 1022 SEFGHAGAKTTTMSLJGLFPPTSGTAYIHDIRTEDEIRKMLGMCPOHNVLEMLT 1081
QY 990 VEEHMFYARLKLSEKHYKAMEOMALDVLGKPSKLSKTSQLSGOKRKSTVALAVG 1049
Db 1082 VEEHMFYARLKLSEKHYKAMEOMALDVLGKPSKLSKTSQLSGOKRKSTVALAVG 1140
QY 1050 GSKVYVLDLPTAGVDPYSRGIMWELLKRYRGRITILSTHMDVADVLGDRITASHGL 1109
Db 1141 GSRAILIDLPTAGVDPYARARIMDLILKYPGRITILSTHMDVADVLGDRITASHGL 1200
QY 1110 CCGVSSFLKNOGLGYVYLLVKKDVESSLSCRSSSTVSLTKEDSVSSSSAGIGS 1169
Db 1201 KCCGSPFLKGYIGDYRLTLVKKRAEPG-----GPOEPGLAS 1238
QY 1170 DHESDTLTIDVS--AISNLIRKVSSEARLVEDIGHELTYVLYPEAKEGAFVLEHEDD 1227
Db 1239 SPGRAPLSSCSELQVSGQIRKHVASCLLVSTSTELSYILPSEAKAGCAERLFOHLER 1298
QY 1228 RLSDLGISYSGSETTLEELFLKVAE-----SGVDA-ETSDGTLP----- 1267
Db 1299 SUDALHLSFGLMDTTLLEVLFLKVSSEDOSLENSADVKESSKDVLPQAEGPASGEHAG 1358
QY 1268 -----ARRNRR--FGDRQSLRPFTEDDADPNDSDIDPESR 1303
Db 1359 NLARSELTOQASLOASVSGSARGDGAGTYGYRPLF-DNODDP--NVSLOEV 1415
QY 1304 EFTDLISGMDKGSYOVKWKMLTQOQFVALLKRLILARSRKGFPAOTVLEPAVECIALV 1363
Db 1416 EAEALSRV-GQGSRLIDGWLKVRQFHLVLRFRHARNSKALFSQILLPAFVCAVMT 1474
QY 1364 FSLVPPGKVPSTLQWMMYNEQY-----FVSDAE-----DTGTELLNAL 1408
Db 1475 VALSPETIGDLPVLVSPSYH-NITOPRGNFIPYANERREYRLRLSPDASPQOVLSTF 1533
QY 1409 TKDPGFTGRM-----EGNPI----- 1424
Db 1534 RLPSGVATCVLKPSPANGSLPPTLNLSSGESRLAARFEDSMCLSTFQGIPLSNFVPP 1593

```

```

QY 1425 -----PD-----TPQAGEEBWTAP-VPOTIMDLFONGMNTQONPSPAC 1463
Db 1594 PSPAPSDPAPDEDLQAMWNYSILPTAGREEMTISAPSLPRVREPRV-----C 1641
QY 1464 QCSSDKIKKMLPVCPGACGLPPPOKRONADLLODLTGRNISDYLVKTYVOJIAKSLN 1523
Db 1642 TCSAGTGFCS--CPSSVG--HPQMRVVTGDLITDTIGHVSEYLLFTSDRF----- 1690
QY 1524 KIWVEFFRYGGFSLGVSNTQALPSSQEVNDAIKMKHKLKLANDSSADRNLISLGRMTG 1583
Db 1691 -----RLHRYGALITG--NLKSTPASTCTRAPRVRK----- 1721
QY 1584 LDTNNKVKWENKNGMAISSFLVNNAILRANLOKE-NPSHYGITAEHNPJNLTKQO 1642
Db 1722 IAVBRAQVFNKNGYHSMPTLYNSLNNAILRANLPRSKGNPAAYGTTVNNHMMNTSAS 1781
QY 1643 LSEVALMTTSDVLYVSLCVIFAMSFVASTVVFLLQERVSAAKHLQISGVKPVYIWSN 1702
Db 1782 LS-LDYLLQGTDDVIAIFIIIVAMSFVASEVFLVLAESTKAKHLOFVSGCNDIYWMAN 1840
QY 1703 FVMDMNVVVPATVITIFLFCFOOKSVSSNLPVALLLLLYGMSTPLMYPASFVKI 1762
Db 1841 YVMDMLNTVPAICCVIILVFDLPATTPNPPAVSLFLLGWSTTPIMYPASFVEX 1900
QY 1763 PSTAYVVLTSYNEFIGINGSVAFVLELT-DNKLNNINDILKSVPLIPHPHICRGLID 1821
Db 1901 PSSAVVELIVINLEFIGITAVATFLLQLFEDHDKLVNSYLSKSCFLPNNYLGHLME 1960
QY 1822 MVRQAMADALERGE-NREVPISMDVGRNLFAMAABEVFELTYVLIQYFEFLRPR 1880
Db 1961 MAYEYINETYAKIGQFDKMSPEMDIVTRGLVMAABEVGFLITIMQYVFLRPR 2020
QY 1881 VNAKLSPLNDEDEDVREBORILDGGGONDLIEIKELTKIYRKK---RKPVRVICGIP 1937
Db 2021 MPVSTKRVED-DVDVAERQRLVLRGADNDNMVAINENLTKYKSKRIKGRILLADRLCLAR 2079
QY 1938 PGCEFGLLGVNAGKSSFTKMLTGDTTVTRGDAFLKNSITLSINIEHONMGYCPQFAD 1997
Db 2080 PGCEFGLLGVNAGKSTFKMLTGDESTTGGEAFVNGHSHVLKELILOOOSLGCPCQDAL 2139
QY 1998 TELLTGREHEFPALLGVGEKEVGAIRKLGIVKGEYVAGNYSGNKKRKSTAM 2057
Db 2140 FDELTAREHELOITRLRGISWKDEARVYKALKELELTADKPACTYGGCKRKRLSTAI 2199
QY 2058 ALIGPPVFLDEPTGMDKARFLMNCALSVYKESGVYLTSHSMECEALCTRMATM 2117
Db 2200 ALIGPPAFILDEPTGMDKARRFLMNLIDILIKTGRSVLTSHSMECEALCTRLAIM 2259
QY 2118 VNGFRCLGSVOHLKRNFGGYTIVRILAGSNPDLPYODFGCLAPPGSVLKEKRNMLQ 2177
Db 2260 VNGRLKLGSTIOHLKRNFGGYTIVR-TRSSQSVADVRFERNRNPDEAMLERHHTKVO 2318
QY 2178 YOLPSSLSLARIETLSQSKRRLIEDYSVSQTLIDQVFVNFKAQSDQD 2227
Db 2319 YOLKSEHISLAOVFSKMEQVSVGLJIEDYSVSQTLIDVNFVNAKQSDN 2368

RESULT 13
US-10-072-621-8
; Sequence 8, Application US/10072621
; Patent No. US20020169137A1
; GENERAL INFORMATION:
; APPLICANT: Relinet, Peter B.
; APPLICANT: Comop, Bruce P.
; TITLE OF INVENTION: REGULATION OF AMYLOID PRECURSOR PROTEIN EXPRESSION
; FILE REFERENCE: 100103.402
; CURRENT APPLICATION NUMBER: US/10/072,621
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 8

```



```

? LENGTH: 2001
? TYPE: PRT
? ORGANISM: Homo sapiens
? FEATURE:
? NAME/KEY: VARIANT
? LOCATION: 30, 70, 280, 477, 558, 1471, 1651, 1689, 1724
? OTHER INFORMATION: xaa = Any Amino Acid
? FEATURE:
? NAME/KEY: VARIANT
? LOCATION: 30, 70, 280, 477, 558, 1471, 1651, 1689, 1724
? OTHER INFORMATION: xaa = Any Amino Acid
US-10-072-621-8

```

Query Match	34.28;	Score 4030;	DB 9;	Length 2001;
Best Local Similarity	43.88;	Pred. No. 5.2e-311;		
Matches 886;	Conservative 294;	Mismatches 549;	Indels 296;	Gaps 46

OY	382	KILYRDRPATGQVAENKTPFOELAVPHDELGMEELSPKWTFMENSOQMDLVRLD	441
Db	26	KILYRPAQSEVDVLLKNEFPAFGVNTYHAQWLNTSAEIRSLXLDQROOLRWL	83
OY	442	SRNDHFEQOOLDLMDTADIVAFIAKHPEDYOSSNGSVYTWSEAFENFN	492
Db	84	-----QQYVALERHRE---ALNLSIDELPRALROONFLPSGSMAL	121
OY	493	-QAITIS-----RFEQCNLKLBERLATEWLINXKSE--LLDERKFWAGIVFTGIT	542
Db	122	LQOOLDTDMAPCSGIOENKSVYDIFGFGFDEESIVYNTLNOAYODVNTYFACVIEQTRK	181
OY	543	PGSLDPHNHYUKITIMODNVERNKIKQDQWDRGPADRFEDMKRYWGGFAVLQDREY	602
Db	182	DGS--LRPHVHYKIRONSFTFEKINELIRAWHGRPTG---GRFYLFGFWYIDMMER	236
OY	603	ALIRVLGTE-KKQGVYMOQPRCYVDILFLRYMSSMBELMTLAWISAVYIKGIYV	661
Db	237	ALIDTFVCHDVVEGSGYVOMPRYPCYRDDFLFVIEHMMOLCXHISVYSVANTQHIYA	296
OY	662	EKEARKFKTRIMGIDNSILMFSEFISLLRPLYSAGLLVLIATKGNLPRDPSVYFV	721
Db	297	EKEHRLKRYMKTMLNNVHNVAFITGFVQVLSISTVYALAIILKYGOVLMHSIVYIWF	356
OY	722	LSFPAVNTILOCFLISTLFSRANLAAAGGIIYFTLPLVPLC-----VAMQDYGFYTK	776
Db	357	LAVYAVNTIMCFVLVSUYLSKAKLASA-GGIYFLVSVPYMYVALREYVAHDKITAFE-K	414
OY	777	IFASILSVPAQECCEAFALFEEDGIGVQNDNLFESVEDGNGNLTTISMLFDTFLYG	836
Db	415	CIASIMSTAGLSKSKYFALEYAVAGVSIQHTGESQSEDEDDNLLAVYMLVWDVAVVG	474
OY	837	VMTYIEAVFPGQYGIQPRWYFPCTKSYMGE---EEDKSHSGSNOKRMS-----EIC-	887
Db	475	ILXMYIEAVHNGMYGLPRWYFPLQKRYMLSGRTLEMENSMWAKIPLUSVMEEDQACA	534
OY	888	-----MEEBPHLKLGVSIONLVYVRDGMKAVAYDGLALNFYEGQITSPFLGHNG	936
Db	535	MESRREFEETRMEEPHLEPLVUXVYDCLTYVYKDDKRLALNLSLMIYENOGVSFLGHNG	594
OY	937	AGKTTTMSILTGLEPPPTSGFAYILGKQIREBMTIQNLGVCQGHNVLFDMLTVEHIMF	996
Db	595	AGKTTTMSILTGLEPPPTSGSATIYGHDIRTEMEIKRN-GHNPQHNVLDFDLRLEHIMF	653
OY	997	YARLGLSEKHVKAEMQOMALDVGRLPSSKLSKTSQLSGGMOKRISVALAFPGSGSVTL	1054
Db	654	YSRLKSAQOEEIPEEMDKMIEDLEL-SNKRHSLVQYLSGGMKRYKVSVALAFPGGSAITL	712
OY	1057	DEPRAGVDPYRSRGVIMELLKTYKQGRITIIISTHIMDEADVLDGRALAIISHGKLCVGS	1114
Db	713	DEPRAGVDPYARRAIRIMOLIKYRPGRTILSTHIMDEADLLGRALIIISHGKLCVGS	772
OY	1117	FLKNQLOSTGYUULLVKKYDVSSLSISCRNSSSVSYLKKEDSVSSSDGAGLSCHSEDTL	1176
Db	773	FLKQYDGDGRILTLVVRPARG-----GPQEGGLASSPPGRAP	810

Qy	1177	TIDVDS--ATSNIILRKHVSARIVMEDIGHETLTVLPYPAAREGAFVFLFHEIDRLSDLCI	1235
Db	811	LSSCELOVSOFTKRKHVNSCLVSTSTELSTILPSPAAKGAERFLFOHLERSIDLHL	870
Qy	1235	SSYGISETTLLEELFLKVA-----EESGVDA--ETSODTLKARRNR-----	1272
Db	871	SSFGMLDITLLEEVFLKVSGGDSSLSENSADVKESERKIVLPGAGHSGBEHGNLARCE	930
Qy	1273	^-----RAFGDKQSCILRPETE-----DDAADPNDSDIDPESRETDIL	1306
Db	931	LTIQSQASLOSASSVGSALGDGGA---GYTDVGYDPELPFNPDOPD--NVSLOEVEARAL	985
Qy	1309	SGMDGSGSYQVSGMKLTDOOFALLMLMKRLIARSRSKGFPAIIVLPAVVCIALVPSLIV	1368
Db	986	SRV--CGGSKRLIDGMLKRYKOFHGHLLVRFHCARNSKALFSQLILPAFFVCYAAIYVALSV	1044
Qy	1369	PPPGKTPSLELOPMYINBOYT-----FVSNDAPE-----DTGTELLMLNATKDPG	1413
Db	1045	PETGDLPLPLVLSPOYH--NYTOPRGNFIPYANEEERREYRLRLSPDASPOOLVSTFRLPDG	1103
Qy	1414	FGTRCM-----EGNPI-----	1424
Db	1104	VGATCYLSPANGSLGPTLNLNSESERLLAARFPDSCLESFTQGLPLSNEYPPRPPSPAR	1163
Qy	1425	-----PD-----TPCAGEEETAP--VPOTIMDLFQNGNMTMONPSPACQSSD	1468
Db	1164	SDSPASPDDEDLOAMVNSLPPTAGCEMWTSPASPLRLVREPV-----CTCSAQ	1211
Qy	1469	KIKKMLPVCPPGAGLPPPOKKONTAIILODLGRNISDYLYKTYQIIAKSLKKIWN	1528
Db	1212	GGGFS-----CPNSVVG--HPQOKRVVATGSLIDTIGHVSEVLETTSDRF-----R	1256
Qy	1529	FFRYGGSILGVSNTQALPPSOEVNDALIKQKKHLIKAKSSADRFPLNSLGRFMTGLDTR	1588
Db	1257	LHRHAYATFGG--NVILKSIPAS--FCTRAPVVRKIRICRAA-----	1293
Qy	1589	NKXVFNFKGWHAISSPLNTVNNAILRANLOKKE--NPSHYGTAENHPLNTKOOLESVA	1647
Db	1294	--OVEYNNKGHSMPTJYINSJLNNAILRANLPKSKGPAAVGITVNNHPNNKTSASTLS--LD	1350
Qy	1648	LMTTSVDVLVSICVLFAMSEVPASVFFVFLLOEVSRSKAKHLOFISGKRPITWLSNFWDM	1707
Db	1351	YLLGQDVVIAIFLIVAMSFPASFFVFLVLEAKSTAKHLOEYSGCNPITLIVLANYVDM	1410
Qy	1708	CWVVPATLVIITFLPCQOQSYSSVSNLPVLLALLLLYQWSTPPLMYPASVFKIPSTAY	1767
Db	1411	LWYLVAPPCCVIILFVDPDLATYSPTNFPNAVLSLFLYQWSTPPLMYPASVFEVPSAY	1470
Qy	1768	VVLTVSNLFETIGINGSVAFVFLDELFT--DNKLNNINDILKSVLPIPHFCRGRLIDMWYNO	1826
Db	1471	XPLIIVINLFITITATVATFLLOLPEHDKDLKVNYSLSKSCFLPFPVNLGHOLEMAYNE	1530
Qy	1827	AMADALERFGE--NRFPVSPLSNDVLGRNLFAVAEYVFLITVLIOYRFRTIPREPVNKL	1885
Db	1531	YINEYFAKIQCFDMMKSPFEDWITVGRGLVMAVEGVGFLTIMQYNFLRRPQMPPYST	1580
Qy	1886	SPUNEDEDVAREKORILLDGGONDILLEIKELTKYTRK---RKAVNDICVGIPIPEGCF	1942
Db	1591	KPEVD--DVDVASEQORVLRGADNDIMKLEINLTKYVKSRIGRILLAVBRLCIGVAPGECF	1649
Qy	1943	GILLGVNAGKSSFRKMLTGCTVTVRGDAFLNKSJLSJNHEVQNMNGVCPORDATTELLT	2002
Db	1650	GXLGVNAGKSTFKMLTGDESTTGGAFVNGHVSJLKLXQYQOGLSTGCPDODALFLBELT	1709
Qy	2003	GREHVEFALLRGVPEKEVEKVCGBMALIKLIGLVKYGKAGVNSGCKNRKLSLTAALIGC	2062
Db	1710	AREHLOITLTRLGIXMDEARVYVWMLKLELTKYADKPAGCYVSGCKNRKLSLTAALIGY	1769
Qy	2063	PPVVEFLDEPTTGMPKARRLLMNCALSVYKEGSRVLTSHSMECEALCTPMALINVGRF	2122
Db	1770	PAFIFLDEPTTGMPKARRLLMNLIDILKTRGSVLTSHSMECEALCTPLAIVWGRL	1829
Qy	2123	RCLOSVQHLKNRGDDGTTIVVRIAGSNPDLPVQDFGLAPPGVSLKEKRNMLQYOUPS	2182

Dd	1830	RCLSTOHLKRRFEDGIMLIVR-TRKSQSKYDVRFRFNPNRFEMAKLEHNHKKVQYLQS	1888
Oy	2183	SLSLSIARIFSLISQSKRLIEDYSVSQTLLDOVFVNPAKSDSD	2227
Dd	1889	EHIISLAQVFSKMEQSVGLIEDYSVSQTLIDNVFNPAKKQSDN	1933

```

RESULT 14
US-09-995-542-8
: Sequence 8, Application US/09995542
: Patent No. US20020127647A1
: GENERAL INFORMATION:
: APPLICANT: Shuttler, John
: APPLICANT: Ullas, Iainn
: TITLE OF INVENTION: ATP-Binding Cassette Transporter-Like Molecules and
: TITLE OF INVENTION: Uses Thereof
: FILE REFERENCE: 00-658-A
: CURRENT APPLICATION NUMBER: US/09/995,542
: CURRENT FILING DATE: 2001-11-28
: PRIOR APPLICATION NUMBER: 60/253,520
: PRIOR FILING DATE: 2000-11-28
: NUMBER OF SEQ. ID NOS: 24
: SOFTWARE: PatentIn Ver. 2.0
: SEQ. ID NO. 8
: LENGTH: 1550
: TYPE: prt
: ORGANISM: Homo sapiens
US-09-995-542-8

```

Query Match	32.5%;	Score 3832;	DB 10;	Length 1550;
Best Local Similarity	50.0%;	Pred. No. 1.9e-295;		
Matches 762;	Conservative 259;	Mismatches 430;	Indels 74;	Gaps 18

[illegible]

Qy	829	LDFDFLTKWMTWYIEAVFPEQOYIIPRPWYPCMKSWPCEDEKSHGSMOKRMSEJCM	888
Dd	600	LDDALGLTWTWIEAVCPQOYIIPRPWYPCMKSWPCEDEKSHGSMOKRMSEJCM	658
Qy	889	EEEPJHLKLGAVSIQNLVKVYRDGMKVAVDGLALNFEQOITSFLGHNAGKTTMSILTG	948
Dd	659	EEAPRGLSPGVSRLKRPQSPQALRGSLIDFQOHIITAFGLHNGAGTWTLSISG	718
Qy	949	LEPPTSGATILCKDIRSEMSITRQMLGCPQONVYFQMLVYEEHWPYARKJGISEKHV	1008
Dd	719	LEPPSGASFLIHDVHSSAAALRPLHVCQPOINVLFDMLYDEHMYTGRKIGLSAAVY	778
Qy	1009	KAEEMOALBYDJPSSKRLSKTQSLQSGMOKRLISVALAFVGSKVYILDEPTAGVPSYR	1068
Dd	779	GPEODRILQDGL-VSKSQOYQTRHLSGMOGRKLSVALAFVGSQVYILDEPTAGVAPSR	837
Qy	1069	RGIMELLKTRQCKTILSLTHHMDADYADGRATISHRKLCVSSSLFKMQJSTGYL	1128
Dd	838	RGIMELLKTRQCKTILSLTHHDEDELLGDRVAVAGRLCCQSPFLFRHRLHSGYLL	897
Qy	1129	TLVKVDVSSLSSCRNSSYVYLKKEDEVSSQSSDAGLSGSHESPDTLIDVSAISNLR	1188
Dd	898	TLVKARLPDLTN-----EKADTMESGVDTQREKKNGSGHYCPQALLAYQ	945
Qy	1189	KHNSBARLVEDIGHETLYVLYPEAKKGAVALFHEIDRLSDLGSSGISEETLEIF	1248
Dd	946	HNVGARLVELELPHLELVLVLYPGAMHGSFATLFRLELDRLAERLTGYSIDSTLEIF	1005
Qy	1249	LKVAEESVDNETSDGTLPARNRNRAFQ--KQSCLEPTEEDAAPRNDSDIDPESRETL	1307
Dd	1006	LKVVECAADLDMDEGSCGGLHLCIGLADLVTLRLKMPQETALE--NCPAGSAPETQD	106
Qy	1308	LSGMDKGSYOVKCMKLTQOQFVALMKRLILRRSRKGFPAOYLYPANYALVLSLI	136
Dd	1064	GSGRPAVQ--EYQGMALLRQDOLALLKRLRLARRRGRGLFAOYLYPALFVGLAVFSLI	112
Qy	1366	VPEPKYPSLELOPMWYNEQYTFVNSAPBDTQLELNLATIDPQFGTRCMGDNIPDT	1427
Dd	1122	VPEPKYPSLELOPMWYNEQYTFVNSAPBDTQLELNLATIDPQFGTRCMGDNIPDT	117
Qy	1428	PCQAGEEWTAPYRQYIMDLFQNGNTMONPSPACQSSDKTKKMLPVCPRGAGLPP	148
Dd	1172	PYQHSRHRSAPRYPAEYAKVLASGNWTPSPSPACQSSQPGARRLLPDCPPAAAGPRP	123
Qy	1488	QKKQMTAIIODLGRNLSIDVLYTYOYLAKSKKIMWNERRYGGSFLGYSNFOALP	154
Dd	1232	QAVYTSGEVONLGRNLSIDVLYTYOYLAKSKKIMWNERRYGGSFLGYSNFOALP	129
Qy	1548	SOEVNDAIKQMKHLKLAKDSSADRFNLISGREFMGDLTRNNVYKVPNNKGMAHAISSPLN	1607
Dd	1291	GOELGRSVEELMALISPLPGGALDRVLRKNTLVAHNSIDADSCLKIMFNKNGHMSHVAFN	1350
Qy	1608	VYNNALILANLOKQENPSHYGTFPNRPLNLTQOJSEVALMTYSDVLYSICVIFAMSF	1667
Dd	1351	RASNAILRNLHLPGRFARHNSITTLNHRPLNLTQJSEALMASSVDVLYSICVIFAMSF	1410
Qy	1668	VPAFVEVLIQERYSKAKHLOFISGVKPVYWLNSFVDMQCNVBDATVLIIFLICPOQ	1727
Dd	1411	VPAFVLIYIEBRYBRAKHLOLMGSLSPLYWLGNFLMDQCNLYPACIVLIIFLAFQOR	1476
Qy	1728	SVYSSNLPYVALLLLLYGMSITPL 1752	
Dd	1471	AVVAPRNPALLLELLYGRQAVPY 1495	

```

RESULT 15
US-09-767-870-9
; Sequence 9, Application US/09767870
; Patent No. US20020037549A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: ABC Transport Polynucleotides, Polypeptides, and Antibodies

```

Fri Mar 7 09:16:52 2003

us-09-595-526c-2.rapb

FILE REFERENCE: P7010P1
CURRENT APPLICATION NUMBER: US/09/767,870
PRIOR FILING DATE: 2001-01-24
PRIOR APPLICATION NUMBER: 2000-07-20
PRIOR FILING DATE: 1999-07-23
PRIOR APPLICATION NUMBER: 60/145,215
PRIOR FILING DATE: 1999-08-18
NUMBER OF SEQ ID NOS: 60/164,730
SOFTWARE OF SEQ ID NOS: 1999-11-12
SEQ ID NO 9
LENGTH: 664
TYPE: PRT
ORGANISM: Homo sapiens
US-09-767-870-9

Job time : 97 secs

Query Match
Best Local
Matches 397: Similarity 17.4% Score 2054: DB 10: length 664:
DB 1604 SPINNYNNALIRANLONKGENPSHYGTAFNHPNLTKOOLSEVALMTTSDVLSYCF 1663
OY 3 AFVNRASMAILRAHLPBGPBARHANSITTELNHPLNTRKEOLSEGALMASSVDLYSCVIF 1663
DB 1664 AMSEVPASVVELLOEYKSKAKHLOFISGVPRVYIMLSNFWDMCNVYVPAITLYIIFC 1723
OY 63 AMSFVPASETLVILERRVYRAHNLQMGSLPPLYWLGFMDCNTLVPACTIVLFLA 122
DB 1724 FOOKSYSSSTNLPVALLLLLYGKSTPLMYPASVFKRIPSTAYVYVSVMLFIGINSV 1783
OY 1784 AFVLELPTDNKLNINDILKSVFLFPPHCLGRLGIDMKNOMADALERGENREVS 182
DB 183 AFVLELPTDNKLNINDILKSVFLFPPHCLGRLGIDMKNOMADALERGENREVS 182
OY 1844 LMSDLVGRNLPMAVEVEVEFLITVLYOYRFTIRPPVNAKLSPLNDEDEDVARERORIT 1903
DB 243 LMEVEYKKNLAAVIGOGPLFLFTLLLOHRSOLLPQPRVBSLPLGEEDEDAVARERERV 302
OY 1904 DSGGNDILEIKELTKYRRKRAVDRIYCIIPGCEFGILGVNGAGKSTFRMLTGD 1963
DB 303 OGATOGDVLVRLNLTLYYRGORAPVADRLCLGIPGCEFGILGVNGAGKSTFRMLTGD 1963
OY 1964 TVTGGDAFLKNSIISNIEHYHOMKGYCPOPATITELTGREHYEFFALLRGVPEKEVGK 2023
DB 2024 VGMALIRKLGIVYKGYAGNYSNGNKRKLTSTMALIGRPVVELDEPTTGMDPKARRFL 2083
OY 2084 WNCALSVYKGRSVYVLSHMECEALCTRAIMVNGRFRCLGSVOHLKNRFGDGTIV 2143
DB 483 RINGSNBDLAKPVODFFGLAPGSLYKREKRMLOYOL-PSSLSLARIFSTLSOSKRRLH 2202
OY 2203 IEDSVSOTLQOVENVNFAKKOSDDHLKDLKLNQTV-VDAV-----LTSFLQDE 2254
DB 601 VEDFSVSOTMLEEVLFSKDGKDE--DTEBKEAGVGVDAAGLOHPRKVSOFELDDP 657
OY 2255 KYKES 2259
DB 658 STAFF 662

Search completed: March 7, 2003, 08:58:22

